



AG Optimierung von Erhebungsdesigns bei Mixed Mode



# Evaluierung und Dokumentation der Rekrutierungsexperimente bei der PUMA-Erhebung Q2/2016

Im Auftrag von der Plattform für Umfragen, Methoden und empirische Analysen (PUMA) und Statistik Austria

erstellt von  
Dr. Alexander SEYMER

Letzte Revision: 17. Januar 2017

---

Ich möchte mich bei der Statistik Austria und im Besonderen bei Matthias Till für die ausgezeichnete Zusammenarbeit bedanken. Die ausführlichen Diskussionen haben wesentlich zur Verbesserung und Schärfung des Berichts beigetragen. Außerdem gilt mein Dank Martin Weichbold für seinen Rat und die Hilfe bezüglich verschiedener Detailfragen.

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>2</b>
<b>Tabellenverzeichnis</b>	<b>3</b>
<b>1 Einleitung</b>	<b>5</b>
<b>2 Incentives</b>	<b>6</b>
<b>3 Webbefragung</b>	<b>7</b>
<b>4 Experiment</b>	<b>9</b>
<b>5 Datenqualität</b>	<b>11</b>
<b>6 Evaluationskriterien</b>	<b>14</b>
6.1 Ausschöpfung . . . . .	17
6.2 R-Indikator . . . . .	18
6.3 Effektive Stichprobengröße . . . . .	20
6.4 Erhebungskosten . . . . .	21
6.5 Variablen für die Untersuchung von Selektivitätseffekten . . . . .	21
<b>7 Evaluation</b>	<b>23</b>
7.1 Stichprobenfehler . . . . .	23
7.2 Non-Response-Fehler . . . . .	24
7.3 Gesamtfehler über das Rekrutierungsexperiment . . . . .	31
7.4 Anpassungsfehler . . . . .	32
7.5 Kosten . . . . .	33
<b>8 Schlussfolgerungen und Empfehlungen</b>	<b>34</b>
<b>9 Literaturverzeichnis</b>	<b>37</b>
<b>10 Anhang</b>	<b>39</b>
10.1 Kodierung der Variablen . . . . .	39
10.2 Kreuztabellen zum Vergleich Teilstichprobe und Bruttostichprobe . . . . .	44
10.3 Kreuztabellen zum Vergleich Teilstichprobe und Nettostichprobe . . . . .	51
10.4 Kreuztabellen zum Vergleich Brutto- und Nettostichprobe . . . . .	58
10.5 Ergebnisse logistische Regression . . . . .	66

## Abbildungsverzeichnis

1	Survey life cycle from a quality perspective . . . . .	13
2	Repräsentativität über die Stichprobenziehung im Rekrutierungsexperiment . . . . .	14
3	Odds Ratios aus der logistischen Regression auf die Auswahl aus der Teilstichprobe MZ (Rotation 50) in die Bruttostichprobe . . . . .	24
4	Balkendiagramm zur Ausschöpfung über Antwortkategorien für Bildung nach Incentivegruppen . . . . .	27
5	Balkendiagramm zur Ausschöpfung über Antwortkategorien für Altersgruppen nach Incentivegruppen . . . . .	28
6	Balkendiagramm zur Ausschöpfung über Antwortkategorien für Beruf nach Incentivegruppen . . . . .	28
7	Balkendiagramm zur Ausschöpfung über Antwortkategorien für Einkommensquintile nach Incentivegruppen . . . . .	29
8	Odds Ratios aus der logistischen Regression auf die Auswahl aus der Teilstichprobe in die Nettostichprobe . . . . .	32

## Tabellenverzeichnis

1	Vollständigkeit nach der Anzahl an beantworteten Fragen und dem Grad der Vollständigkeit der Fragebogenbeantwortung . . . . .	9
2	Rekrutierung aus der Bruttostichprobe nach Incentives . . . . .	11
3	Überblick Evaluationskriterien . . . . .	22
4	Ausschöpfungsquote über die Incentivegruppen im Rekrutierungsexperiment . . . . .	25
5	Logistische Regressionskoeffizienten der Stichprobenausschöpfung (Von Brutto- zu rekturierter und Nettostichprobe, Standardfehler in Klammern). . . . .	26
6	R-Indikator für die Incentivegruppen über den Verlauf des Rekrutierungsexperimentes . . . . .	30
7	R-Indikator und Pseudo- $R^2$ der logistischen Regression für Ausschöpfung aus Brutto- zu Nettostichprobe nach Incentivegruppe . . . . .	31
8	Gewichtete Mittelwerte, Standardfehler und Designeffekte nach Lumley einzelner Variablen aus der PUMA Befragung . . . . .	33
9	Notwendige Bruttostichprobe bzw. rekrutierte Stichprobe nach Incentivegruppe unter Berücksichtigung der Ausschöpfungsquoten um eine effektive Nettostichprobe von 1000 zu erreichen. . . . .	33
10	Kostenschätzung basierend auf der ermittelten notwendigen Bruttostichprobe . . . . .	34
11	Dokumentation zur Umkodierung der ursprünglichen Variablen. . . . .	40
12	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Incentivegruppen (ungewichtet) . . . . .	44
13	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Geschlecht (ungewichtet) . . . . .	44
14	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Altersgruppen (ungewichtet) . . . . .	45
15	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Familienstand (ungewichtet) . . . . .	45
16	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Haushaltsgröße (ungewichtet) . . . . .	46
17	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Bildung (ungewichtet) . . . . .	47
18	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Lebensunterhalt (ungewichtet) . . . . .	48
19	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach beruflicher Stellung (ungewichtet) . . . . .	48
20	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Staatsangehörigkeit (ungewichtet) . . . . .	49
21	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Herkunft (ungewichtet) . . . . .	49
22	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Einkommensquintilen (ungewichtet) . . . . .	50

23	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Incentivegruppen (ungewichtet) . . . . .	51
24	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Geschlecht (ungewichtet) . . . . .	51
25	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Altersgruppen (ungewichtet) . . . . .	52
26	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Familienstand (ungewichtet) . . . . .	52
27	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Haushaltsgröße (ungewichtet) . . . . .	53
28	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Bildung (ungewichtet) . . . . .	54
29	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Lebensunterhalt (ungewichtet) . . . . .	55
30	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach beruflicher Stellung (ungewichtet) . . . . .	55
31	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Staatsangehörigkeit (ungewichtet) . . . . .	56
32	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Herkunft (ungewichtet) . . . . .	56
33	Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Einkommensquintilen (ungewichtet) . . . . .	57
34	Vergleich Brutto- und Nettostichprobe nach Incentivegruppen (ungewichtet) . . . . .	58
35	Vergleich Brutto- und Nettostichprobe nach Geschlecht (ungewichtet)	58
36	Vergleich Brutto- und Nettostichprobe nach Altersgruppen (ungewichtet)	59
37	Vergleich Brutto- und Nettostichprobe nach Familienstand (ungewichtet)	59
38	Vergleich Brutto- und Nettostichprobe nach Haushaltsgröße (ungewichtet) . . . . .	60
39	Vergleich Brutto- und Nettostichprobe nach Bildung (ungewichtet) .	61
40	Vergleich Brutto- und Nettostichprobe nach Lebensunterhalt (ungewichtet) . . . . .	62
41	Vergleich Brutto- und Nettostichprobe nach beruflicher Stellung (ungewichtet) . . . . .	62
42	Vergleich Brutto- und Nettostichprobe nach Staatsangehörigkeit (ungewichtet) . . . . .	63
43	Vergleich Brutto- und Nettostichprobe nach Herkunft (ungewichtet) .	63
44	Vergleich Brutto- und Nettostichprobe nach Einkommensquintilen (ungewichtet) . . . . .	64
45	Logistische Regressionskoeffizienten der Stichprobenausschöpfung (von Mikro- zu Teil-, von Teil- zu Brutto- und von Teil- zu Nettostichprobe, Standardfehler in Klammern). . . . .	66
46	Logistische Regressionskoeffizienten der Stichprobenausschöpfung von Brutto- zu Nettostichprobe nach Incentivegruppen (Standardfehler in Klammern). . . . .	67

# 1 Einleitung

Im Frühjahr 2016 wurde von der Statistik Austria im Auftrag der Plattform für Umfragen, Methoden und empirische Analysen (PUMA) eine Piloterhebung für zukünftige Sozialwissenschaftliche Umfragen durchgeführt (STAT\_PUMA2016Q2). Dabei wurde auch ein Rekrutierungsexperiment für Webbefragungen<sup>1</sup> durchgeführt. Im Rahmen des Experiments sollte herausgefunden werden, inwieweit Incentives Probleme der Rekrutierung von TeilnehmerInnen an Webbefragungen kompensieren können. In der Literatur sind die positiven Effekte von Incentives auf die Ausschöpfungsquoten<sup>2</sup> für alle Formen der Befragungen gut belegt. Simmons und Wilmot (2004) berichten auf Basis einer Literaturrecherche von Steigerungen von bis zu 7,3 % in verschiedenen Arten von Erhebungen, während Kennedy und Ouimet (2014) für Internetbefragungen eine Steigerung der Ausschöpfungsquote von 24 % auf 36 % belegen. Und Göritz (2006) weist nach der Ausschüttung materieller Incentives eine Erhöhung der Chance um 19 % zugunsten der Teilnahme an Internetbefragungen nach. Dementsprechend stellt sich im Rahmen dieses Berichtes weniger die Frage des grundsätzlich positiven Effekts von Incentives auf die Ausschöpfungsquote, sondern im Mittelpunkt steht die Frage, welches Incentive die besten Ergebnisse bezüglich Kosten und Datenqualität hat, wobei die Ausschöpfungsquote im Rahmen des Berichts als ein Datenqualitätsmerkmal behandelt wird.

Neben der Steigerung der Ausschöpfungsquote wurde untersucht, ob der Einsatz von Incentives Konsequenzen auf die Selektivität oder Repräsentativität hat. Aufgrund des speziellen Designs des PUMA-Experiments ist es möglich, eine differenzierte Analyse der einzelnen Fehler mit Bezug auf die Selektivität und Repräsentativität vorzunehmen. Dabei wird sich zeigen, dass Repräsentativität sehr unterschiedlich operationalisiert werden kann. Eine Einordnung des Experiments kann letztlich nur in einem größeren Rahmen erfolgen und als solches wird die theoretische Diskussion die im Bericht analysierten Bereiche in ein ganzheitliches Konzept der Umfragenqualität einordnen.

Zu Beginn soll aber zunächst ein kurzer Überblick über die verschiedenen Formen der Incentives und die Erfahrungen damit gegeben werden. Im Anschluss wird auf die Besonderheiten der Webbefragung eingegangen, da diese sich von anderen Erhebungsarten unterscheidet. In Abschnitt 4 wird das Experiment mit dem gesamten Design vorgestellt, bevor in Abschnitt 5 über einen ganzheitlichen Ansatz der Datenqualität

---

<sup>1</sup>Es soll hier von Webbefragung in Abgrenzung von Internet- oder Onlinebefragungen gesprochen werden. Als Web (WWW) wird die Darstellung von Inhalten über Webseiten, meistens HTML basiert, verstanden. Das Internet ist die allgemeinere Form der Kommunikation über standardisierte Protokolle, wobei die für Webseiten notwendige Kommunikation über TCP/IP nur eine Form ist. Onlinebefragungen sind noch eine Stufe abstrakter, weil sie jedes elektronische Netzwerk als Medium umfassen, worüber die Informationen zwischen Befragten und InterviewerInnen ausgetauscht werden können (Vgl., Callegaro, Manfreda & Vehovar, 2015, S. 12).

<sup>2</sup>An dieser Stelle sei darauf hingewiesen, dass der Begriff Ausschöpfungsquote im Rahmen des Berichts teilweise weiter verwendet wird als in der Literatur üblich. Ausschöpfungsquote wird meist als das Verhältnis von realisierten zu versuchten Interviews verstanden. Da das Rekrutierungsexperiment die Möglichkeit bietet, verschiedene andere Entscheidungen bei der Stichprobenziehung ähnlich wie die Ausschöpfungsquote zu betrachten, wird dies in Anlehnung so umgesetzt und die Benennung soll die konzeptionelle Nähe reflektieren.

der Rahmen für die Evaluationskriterien definiert wird. Die Kriterien werden dann eingegrenzt detaillierter in Abschnitt 6 vorgestellt. Im Anschluss folgt die Evaluation entlang der vorher festgelegten Kriterien. In den Schlussfolgerungen werden die Ergebnisse zusammengefasst und einige Empfehlungen formuliert.

## 2 Incentives

Incentives können danach unterschieden werden, wann sie vergeben werden und in welcher Form (Vgl. Ernst Stähli & Joye, 2016). Beim Zeitpunkt wird in bedingungslose und konditionale Incentives differenziert. Bedingungslose oder nicht-konditionale Incentives werden beim ersten Kontakt mit der zu befragenden Person unabhängig von der Teilnahme übergeben. Es existieren unterschiedliche theoretische Begründungen für diesen Ansatz. Nach der Theorie des sozialen Austauschs wird das Incentive als Geschenk und eine erste Geste des Austausches definiert, welche die Motivation des Gegenübers zur Teilnahme erhöht. Dabei spielt der Wert des Geschenks normalerweise keine Rolle, denn es geht um den symbolisch initiierten Austausch. Allerdings kann bei diesem Austausch ein zu großes Geschenk eine gegenteilige Wirkung erzielen. Alternativ kann auch mit dem Modell des ökonomischen Austausches argumentiert werden. In diesem Fall entschädigt das Incentive die befragte Person für den Aufwand und ist entsprechend in der Höhe zu wählen. Konditionale Incentives werden zum Zeitpunkt der Teilnahme oder nach Abschluss der Erhebung in Abhängigkeit von der Teilnahme an die befragten Personen ausgegeben. Der größte Vorteil von konditionalen Incentives begründet sich aus den Kosten, denn die Incentives fallen nur für durchgeführte Befragungen an.

Die Unterscheidung nach der Form der Incentives ist komplexer. Zunächst gibt es monetäre Incentives, wobei hier der zu rekrutierenden Person direkt Bargeld übergeben wird. Die Beträge können dabei von wenigen Cent bis zu mehr als 100 EUR betragen. Diese Form ist zwar unmittelbar und leicht nutzbar, allerdings ist die Versendung per Post oder auch die direkte Übergabe durchaus problematisch, da die Befragten irritiert oder misstrauisch darauf reagieren können. Quasi-monetäre Incentives sind Schecks, welche eine Alternative zu Bargeld darstellen, als weniger invasiv betrachtet werden und das Erhebungsinstitut nur bei der Einlösung Geld kosten. Trotzdem sind Schecks unterschiedlich gut in den gesellschaftlichen Subpopulationen vermittelbar und deshalb besteht das Risiko die Stichprobenszusammensetzung zu verzerren (Ernst Stähli & Joye, 2016). Semi-monetäre Incentives sind Gutscheine, welche ähnliche Attribute wie quasi-monetäre Incentive haben, allerdings einen geringeren Einfluss auf die Ausschöpfungsquote zeigen (Ernst Stähli & Joye, 2016). Gutscheine haben außerdem den Nachteil, dass sie nicht universell einsetzbar sind, das Erhebungsinstitut mit einem Wirtschaftsunternehmen in Zusammenhang gebracht wird und auch Kosten bei Nichteinlösen des Gutscheins entstehen können. Nicht-monetäre Incentives können verschiedene Gegenstände wie Stifte, Bücher o.ä. sein, wobei hier ein ähnliches Problem wie mit den Gutscheinen auftritt, nämlich dass diese Gegenstände nicht den gleichen Wert haben für jedes Individuum. Im Allgemeinen sind nicht-monetäre Incentives deshalb auch weniger effizient (Ernst Stähli & Joye, 2016). Zwei besondere Formen von Incentives sind Lotterien und Spenden. Bei Lotterien wird als Incenti-

ve die Teilnahme an einer Lotterie mit einem großen Gewinn ausgegeben, wobei der Anreiz die Lotterie zu gewinnen, der eigentliche Incentive ist. Lotterien sind rechtlich aufwendig bzw. schwierig umsetzbar und haben einen geringen Effekt zur Steigerung der Ausschöpfungsquote gezeigt (Ernst Stähli & Joye, 2016). Die Idee hinter Spenden ist, der befragten Person freizustellen, welche gemeinnützige Vereinigung eine Spende erhalten soll. Diese Form der Incentives hat sich als sehr ineffizient herausgestellt (Ernst Stähli & Joye, 2016).

Die Effekte von Incentives sind während der Rekrutierungsphase zunächst eine Erhöhung der Ausschöpfungsquote. Dieser Effekt ist stärker bei geringen Ausschöpfungsquoten und größer bei einem höheren Aufwand für die befragten Personen. Bei der Rekrutierung kann es auch vorkommen, dass Personen das Incentive als zu invasiv wahrnehmen und die Teilnahme aufgrund des Incentives ablehnen. Dies kann zu Selektivitätsproblemen führen. Mit Bezug auf die Datenqualität findet sich in der Literatur Evidenz für einen positiven Einfluss auf die sozio-demographische Repräsentativität und das Antwortverhalten (Medway & Tourangeau, 2015).

Außerdem müssen auch andere Folgen von Incentives berücksichtigt werden (Vgl. Ernst Stähli & Joye, 2016). In den aktuellen Debatten um Incentives wird oftmals von differenzierten Incentives gesprochen (Vgl. Ernst Stähli & Joye, 2016), um die Incentives fokussiert auf Subpopulationen anzuwenden, wo es größere Probleme mit der Rekrutierung gibt. Diese differenzierten Incentives können von den Anderen, welche an der Befragung ohne oder motiviert durch geringere Incentives teilnehmen, als unfair betrachtet werden. Auch die soziale Erwünschtheit muss insgesamt für alle Formen der Incentives kritisch hinterfragt werden, im Besonderen wenn die Befragung durch öffentliche Gelder finanziert wird. Als Letztes müssen die langfristigen Folgen Berücksichtigung finden, denn die Verwendung von Incentives verändert die Wahrnehmung von Befragungen und hat damit einen starken Einfluss auf zukünftige Befragungen.

### 3 Webbefragung

Webbefragung sind nur eine Form der Befragung. Um die Vor- und Nachteile richtig einzuordnen, sollen hier auch kurz die anderen drei Formen diskutiert werden. Grundsätzlich können die Form der Befragung natürlich sehr stark differenziert werden, an dieser Stelle soll aber eine Unterscheidung in persönliche, telefonische, postalische Befragung und Webbefragung hinreichend sein (für eine detaillierte Diskussionen von Befragungsmodi vgl. Weichbold, Bacher & Wolf, 2009; Reinecke, 2014).

Historisch betrachtet waren (Face-to-face) persönliche Befragungen lange Zeit der Standard. Während der persönliche Befragung kann die befragte Person durch den Interviewer oder die Interviewerin unterstützt werden, komplexere Hilfsmittel können verwendet werden (Visualisierung von Skalen, Fotos, Audio oder jedes andere Objekt) und die InterviewerInnen können trainiert werden, um die Ausschöpfungsquote zu erhöhen. Die wesentlichen Nachteile sind hohe Kosten und Effekte von sozialer Erwünschtheit.

Mit der flächendeckenden Verbreitung von Festnetzanschlüssen wurden Telefonbefragungen attraktiver. Telefoninterviews waren wesentlich kostengünstiger und die



zufällige Auswahl über zufallsgenerierte Telefonnummern ermöglichte einen besseren Zugang zu einer Zufallsauswahl als viele Verfahren bei den persönlichen Interviews. Ein wesentlicher Nachteil der Telefonbefragungen sind die eingeschränkten Möglichkeiten bei der Kommunikation. Visualisierung sind nicht möglich und komplexere Skalen müssen ausschließlich über Audio kommuniziert werden. Das größte Problem von Telefonbefragung ist allerdings die rapide Ersetzung von Festnetzanschlüssen durch Mobiltelefone und alternative Formen der Kommunikation. Heute besitzen viele Haushalte keinen Festnetzanschluss, sondern alle Mitglieder eigene Mobiltelefone bzw. Smartphones, welche zunehmend zur asynchronen Kommunikation (D. Dillman, Smyth & Christian, 2014), wie bspw. SMS, Instant Messaging oder Emails, genutzt werden.

Die postalische Befragung ist bereits so alt wie die persönliche Befragung, wurde aber aufgrund der geringen Ausschöpfungsquoten und der fehlenden Unterstützung der befragten Person lange als qualitativ schwächer eingestuft. Vor allem die fehlende Unterstützung beim selbständigen Ausfüllen beschränkte die Komplexität der Fragebögen. Außerdem war die Coverage ein Problem, wenn keine Informationen zu Adressen der Haushalte vorhanden waren. Ein Vorteil der postalischen Befragung sind die Kosten, da weder Call-Center noch Training von InterviewerInnen notwendig sind. D. Dillman et al. (2014) argumentieren, dass durch die Bereitstellung der Postzustellinformationen des US Postal Service nun eine Abdeckung von 95 % bis 97 % erreicht werden kann und dass postalische Befragungen mittlerweile auch bessere Ausschöpfungsquoten als telefonische Befragungen oder Webbefragungen erreichen. Die Webbefragung ist von den Möglichkeiten im Rahmen der Visualisierung und Bereitstellung komplexer Fragebogendesigns vergleichbar mit persönlichen Befragungen allerdings nach den Anfangsinvestitionen während der Erhebung deutlich preiswerter. Auch die Auswertung ist vereinfacht und die Rücklaufzeit stark verkürzt. Insgesamt vereint die Webbefragung für die ForscherInnen viele sehr nützliche Eigenschaften, allerdings haben Webbefragung ein erhebliches Coverage-Problem. In Subpopulation mit einer hohen Internetkompetenz und -abdeckung, d.h. im Besonderen Personen im erwerbsfähigen Alter und mit Hochschulabschluss, können mittlerweile repräsentative Befragungen basierend auf Webbefragungen vorgenommen werden. Andere Subpopulationen, wie etwa Personen mit niedrigerem Bildungsniveau, niedrigerem Einkommen oder ältere Personen, sind dagegen deutlich unterrepräsentiert (D. A. Dillman & Edwards, 2016; D. Dillman et al., 2014).

Webbefragungen erlauben eine sehr detaillierte Analyse von Non-Response. Neben der typischen Unterscheidung in kein Kontakt, Teilnahmeverweigerung und anderen Gründen (R. M. Groves et al. 2009) unterscheiden Callegaro, Manfreda und Vehovar (2015) zusätzlich noch Item Non-Response, Abbruch und Vollständigkeit.<sup>3</sup> Aus der Tabelle 1 wird besonders die Problematik der Abbrüche im Vergleich zu den abgeschlossenen Fragebögen deutlich. Grundsätzlich dürfte es sehr schwierig sein, abgebrochene Fragebögen in der Nutzbarkeit der Informationen von vollständig durchlaufenen Fragebögen mit fehlenden Antworten zu unterscheiden. Callegaro, Manfreda

---

<sup>3</sup>Eine genauere Einordnung und Charakterisierung des Non-Response-Problems erfolgt im Abschnitt zur Ableitung der Evaluationskriterien. An dieser Stelle soll nur darauf verwiesen sein, dass die von Callegaro, Manfreda und Vehovar skizzierten Dimensionen des Non-Response sehr zentral für die Weiterverwendung der Daten sind.

Tabelle 1: Vollständigkeit nach der Anzahl an beantworteten Fragen und dem Grad der Vollständigkeit der Fragebogenbeantwortung (Vgl. Tabelle 2.2., Callegaro et al., 2015, S. 143)

Number of items responded to	Level of questionnaire completion		
	Introduction breakoff	Questionnaire breakoff	
No item (UCR = 0)	Unusable	Unusable	
Some items ( $0 < \text{UCR} < 1$ )	-	Unusable	Unusable
		Partial*	Partial
		Complete	Complete*
All items (UCR = 1)	-	-	Complete*

Sterne markieren die technische Empfehlung. UCR steht für unit-level completeness rate und entspricht dem Anteilswert der beantworteten Fragen an allen Items im Fragebogen.

und Vehovar verweisen aber darauf, dass die Ausschöpfungsquote nur ein Maß für Datenqualität ist und dass vielmehr der Non-Response-Bias die entscheidende Größe ist. Dementsprechend ist es durchaus plausibel, zunächst die vollständigen Daten bezüglich des Non-Response-Bias zu betrachten und falls hier keine Probleme auftauchen, das Problem des Item-Non-Response konservativer durch Fallausschluss zu handhaben.

## 4 Experiment

Die Basis für die Stichprobenziehung bilden die Mitglieder aus etwa 4.000 Haushalten, welche nach dem 2. Quartal 2016 aus dem Mikrozensus ausgeschieden sind. Der Mikrozensus erhebt quartalsweise Daten von mehr als 20.000 Haushalten in Österreich mit den inhaltlichen Schwerpunkten Erwerbs- und Wohnungsstatistik. Ziel der Erhebung ist die Dauerbeobachtung wesentlicher gesellschaftlicher Veränderungen der Wohnbevölkerung, insbesondere auch der Arbeitslosenquoten im EU-Vergleich. Mit jedem Quartal scheidet ein Fünftel, d.h. rund 4.000 Haushalte, aus dem Mikrozensus aus, und es werden 4.000 neue Haushalte in die Stichprobe aufgenommen. Die Mikrozensus-Erhebung basiert rechtlich auf der Erwerbs- und Wohnungsstatistikverordnung BGBl. II Nr. 111/2010, welche eine Auskunftspflicht der Haushalte umfasst. Die Stichprobenziehung ist nach Bundesländern geschichtet und innerhalb der Bundesländer erfolgt eine disproportionale Auswahl, um auch kleine Bundesländer hinreichend detailliert zu erfassen. In den ausgewählten Privathaushalten werden alle Personen ab 15 Jahren befragt und jeder Haushalt wird maximal fünfmal im Mikrozensus berücksichtigt. Für die Stichprobenziehung des Rekrutierungsexperiments sind deshalb Klumpungs- und Allokationseffekte zu berücksichtigen, denn durch diesen Designeffekt kann die Genauigkeit geringer sein als bei einer einfachen Wahrscheinlichkeitsstichprobe.

Von den 6.388 Personen im Alter zwischen 16 und 74 Jahren, die im 2. Quartal 2016 letztmalig im Mikrozensus befragt wurden, wurden nur die per Telefon (CATI) direktbefragten Personen berücksichtigt. Personen für die andere Haushaltsmitglieder

Auskunft gegeben haben, vor Ort befragt wurden, oder kein Deutsch sprechen, wurden ausgeschlossen. Für dieses Experiment verbleibt somit eine Stichprobengrundlage von 4.249 Personen, welche randomisiert in vier annähernd gleich große Gruppen aufgeteilt wurden. Diesen Personen wurde am Ende des Telefoninterviews folgende Frage vorgelesen, wobei jede Person entsprechend der Randomisierung nur eine Text- bzw. Incentivevariante zu hören bekam:

„Zum Abschluss bittet Sie die Bundesanstalt Statistik Österreich um Mithilfe bei einer Webbefragung für österreichische Universitäten. Es geht um spannende Themen, die die Lebensqualität von allen Menschen in Österreich stark betreffen (wie z.B. Gesundheit und Altersvorsorge).

[4 Textvarianten:]

- A) Als kleines Dankeschön schenkt Ihnen Statistik Austria die neueste Ausgabe von Zahlen-Daten-Fakten, ein Taschenbuch über die Lebensverhältnisse der Menschen in Österreich. Wären Sie bereit, an dieser Internetbefragung teilzunehmen?<sup>4</sup> Diese Gruppe wird im vorliegenden Bericht als Gruppe „Broschüre“ bezeichnet.
- B) Als kleines Dankeschön schenkt Ihnen Statistik Austria die aktuelle 2 Euro-Sondermünze der österreichischen Nationalbank zum Sammeln oder Ausgeben. Wären Sie bereit, an dieser Internetbefragung teilzunehmen?<sup>4</sup> Diese Gruppe wird im vorliegenden Bericht als Gruppe „2 EUR Münze“ bezeichnet.
- C) Als kleines Dankeschön schenkt Ihnen Statistik Austria die aktuelle 5 Euro-Sondermünze mit dem Feldhasen von Albrecht Dürer zum Sammeln oder Ausgeben. Wären Sie bereit, an dieser Internetbefragung teilzunehmen?<sup>4</sup> Diese Gruppe wird im vorliegenden Bericht als Gruppe „5 EUR Münze“ bezeichnet.
- D) Als kleines Dankeschön schenkt Ihnen Statistik Austria einen Gutschein im Wert von 10 Euro, den sie fast überall in Österreich z.B. in Lebensmittelgeschäften einlösen können. Wären Sie bereit, an dieser Internetbefragung teilzunehmen?<sup>4</sup> Diese Gruppe wird im vorliegenden Bericht als Gruppe „Gutschein“ bezeichnet.

Als Antwortkategorien waren folgende fünf definiert:

- Ja, nehme teil (Internetzugang vorhanden)
- Ja, vielleicht (Zusendung näherer Informationen erbeten)
- Ja, bin grundsätzlich interessiert, habe aber kein Internet
- Nein, kann aus anderen Gründen nicht teilnehmen (z.B. Deutschkenntnisse)
- Nein, habe kein Interesse

---

<sup>4</sup>Die InterviewerInnen hatten zusätzlich die Möglichkeit ggfs. den folgenden Text zu ergänzen: „Bevor die Erhebung beginnt, würden wir Ihnen weitere Detailinformationen zusenden (in der Regel in der Folgewoche mit Start 27.04.2016). Das Ausfüllen des Fragebogens dauert nur etwa 5 Minuten. Die Befragung wird einmalig durchgeführt.“

Tabelle 2: Rekrutierung aus der Bruttostichprobe nach Incentives

	Broschüre	2 EUR Münze	5 EUR Münze	Gutschein	Gesamt
Ja, nehme teil (Internetzugang vorhanden)	332	371	409	394	1.506
Ja, vielleicht (Zusendung näherer Informationen erbeten)	12	12	10	8	42
Insgesamt rekrutiert	344	383	419	402	1.548
Ja, bin grundsätzlich interessiert, habe aber kein Internet	24	30	32	27	113
Nein, kann aus anderen Gründen nicht teilnehmen (z.B. Deutsche)	20	30	17	25	92
Nein, habe kein Interesse	727	603	588	578	2.496
Insgesamt	1.115	1.046	1.056	1.032	4.249

Die auskunftsbereiten Personen erhielten im Normalfall binnen 10 Tagen eine postalische Einladung mit Zugangscode zu der Webbefragung und dem entsprechenden Incentive. Zwei Wochen nach dem Avisobrief wurde eine Erinnerungspostkarte übermittelt und alle jene Personen, welche eine E-Mail Adresse angegeben haben, erhielten zusätzlich sowohl nach der Einladung als auch der Erinnerung jeweils 2 Tage später eine elektronische Version der Einladung bzw. Erinnerung. Ergänzend stand den Befragten eine Telefonhotline für Rückfragen zur Verfügung.

Der Webfragebogen bestand aus mehr als 100 Items, wobei die Befragten durch Filterfragen nur einen Teil der Fragen beantworten mussten. Die Filterfragen erhoben die Berufstätigkeit (ja, nein), ob die Person schon einmal berufstätig war (ja, nein) und ob die Person in Pension ist. Darauf abgestimmt wurden Fragen zur Gesundheit und Work-Life-Balance; Arbeit, Alter und Pension; Zukunftsvorstellungen und finanziellen Problemen; zum Alter und gefühlten Alter; zu Steuerfragen und soziodemographischen Charakteristika erhoben (siehe Anhang 10.1 Tabelle 11). Alle Personen mit mehr als 15 fehlenden Antworten wurden ausgeschlossen und 2,7 % (n=113, vgl. Tabelle 2) konnten wegen fehlendem Internetzugang nicht berücksichtigt werden, obwohl die Personen grundsätzlich Interesse bekundeten.

## 5 Datenqualität

Die Qualität von Daten ist mehrdimensional zu verstehen, wobei die verschiedenen Dimensionen teilweise supplementär und teils konträr zueinander in Beziehung stehen. Eine Evaluation erfordert konsequenterweise ein Abwägen, welches sich am konkreten Vorhaben orientiert. Dementsprechend werden in den folgenden Absätzen, ausgehend von einem sehr allgemeinen Rahmen, nur die Aspekte detaillierter diskutiert, welche für das Rekrutierungsexperiment von Relevanz sind.

Eurostat hat im Rahmen seiner Aufgabe der Bereitstellung qualitativ hochwertiger amtlicher Statistiken für die europäischen Institutionen ein detailliertes System an Qualitätsanforderungen ausgearbeitet – den Quality Assurance Framework of the European Statistical System (Committee et al., 2012).<sup>5</sup> Darin werden grundlegende Prinzipien unterschieden, welche sich in drei Bereich gliedern. Die erste Grunddimension umfasst die institutionellen Rahmenbedingungen, welche so gestaltet sein müssen, dass die Effizienz und Glaubwürdigkeit der statistikgenerierenden Institution

<sup>5</sup>Der Quality Assurance Framework of the European Statistical System orientiert sich sehr stark am Konzept der Total Survey Quality (P. Biemer, Trewin, Bergdahl & Japiec, 2014; P. P. Biemer, 2010), welche ähnliche Dimensionen abbildet.

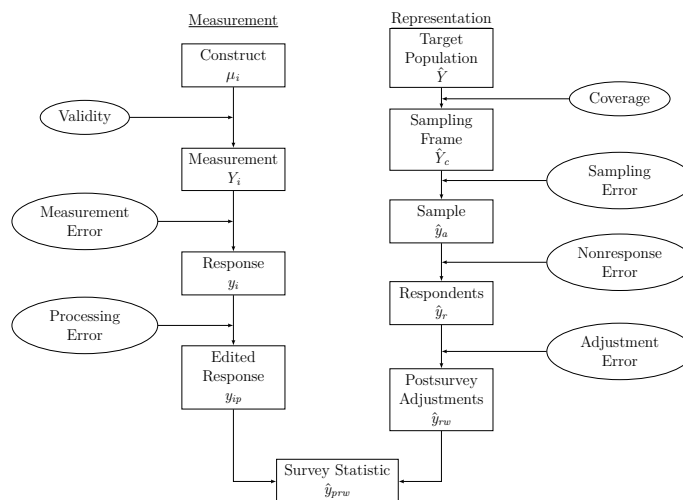
positiv beeinflusst wird. Das berücksichtigt die Verpflichtung zu Qualitätskontrolle, zu Vertraulichkeit im Umgang mit den Daten und zu Objektivität und Unbefangenheit. Diesen Aspekt kann der Evaluationsbericht selbst nicht direkt reflektiert, denn der Evaluationsbericht ist selbst ein Aspekt der Qualitätskontrolle.

Der zweite Bereich berücksichtigt in seiner Gesamtheit den Prozess der Datensammlung, -organisation und -verteilung durch die statistischen Ämter. Dies umfasst eine hochwertige Methodologie, welche die Verwendung der angemessenen Werkzeuge, Techniken und Expertise berücksichtigt, um höchste Standards sicherzustellen. Des Weiteren muss die Datenerhebung und -validierung mit angemessenen statistischen Verfahren umgesetzt werden. Auch die Belastung der befragten Personen muss überwacht werden, damit die Belastung nicht zu groß wird und möglichst über die Zeit abnimmt. Letztlich sind alle Prozesse kosteneffizient zu gestalten, um die Ressourcen zu schonen. Der Evaluationsbericht fokussiert in der Fragestellung auf die Differenzierung der Incentive-Gruppen und weniger auf Aspekte des Designs insgesamt, deshalb werden in die detaillierte Evaluation nur die Kosten und die Wirkung der Incentives auf die Belastung aufgenommen. Allerdings werden in den abschließenden Bemerkungen einige Ausführungen zu den weiterführenden Konsequenzen von Webbefragungen gemacht, um die Einordnung in einen größeren Kontext konsequent weiterzuverfolgen.

Erst die dritte Grunddimension berücksichtigt die statistischen Daten selbst. Bei diesem Qualitätskriterium steht im Mittelpunkt, dass die Daten für die NutzerInnen, seien diese europäische Institutionen, Regierungen, Forschungseinrichtungen oder die Öffentlichkeit, relevant, genau und verlässlich, pünktlich und aktuell, kohärent, vergleichbar über Raum und Zeit und verständlich verfügbar gemacht werden. Unter Relevanz wird kontinuierlich geprüft, ob die Daten den Bedürfnissen der BenutzerInnen gerecht werden. Genauigkeit und Verlässlichkeit soll sicherstellen, dass die Realität nach den höchsten Standards in den Daten reflektiert ist. Dieser Aspekt dürfte für das Rekrutierungsexperiment eine Schlüsseldimension darstellen. Die Pünktlichkeit und Aktualität erfasst die zeitliche Diskrepanz zwischen Beginn der Erhebung und Bereitstellung der Daten, welche möglichst gering zu halten ist. Die Kohärenz und Vergleichbarkeit behandelt die Problematik der Integrität der Daten innerhalb eines Datensatzes aber auch über Datensätze aus verschiedenen geographischen Räumen und Zeitpunkten. Als letztes sind die Daten verständlich aufzubereiten und leicht zugänglich für die NutzerInnen bereitzustellen. Die Kerneigenschaft, welche bei der Evaluation aus dieser Sammlung von Prinzipien als Kriterium übernommen werden kann, ist die Genauigkeit und Verlässlichkeit, denn hiermit ist die Repräsentativität im Kern abgedeckt.

Diese drei Grunddimensionen, Institution, Prozess, Daten stellen einen ganzheitlichen Qualitätsanspruch an Daten und deren Erzeugung und Aufbereitung, welcher weit über die Ansprüche dieser Evaluation hinausgehen. Für ein Rekrutierungsexperiment, wie bereits deutlich geworden sein sollte, sind nur Teile dieser Qualitätsprinzipien relevant. Die Ausführungen sollen aber verdeutlichen, dass die Beurteilung der Qualität von Daten erst in einem Gesamtkontext sinnvoll ist. Innerhalb dieses umfassenden Qualitätsrahmens, stehen in diesem Bericht die Kriterien der Evaluation aus der bisherigen Diskussion auf die Kosten, Wirkung auf die Belastung der befragten Personen und die Genauigkeit und Verlässlichkeit der Daten einschränken im Mittel-

Abbildung 1: Survey life cycle from a quality perspective (Vgl. R. M. Groves et al., 2009, Abb. 2.5, S. 48)

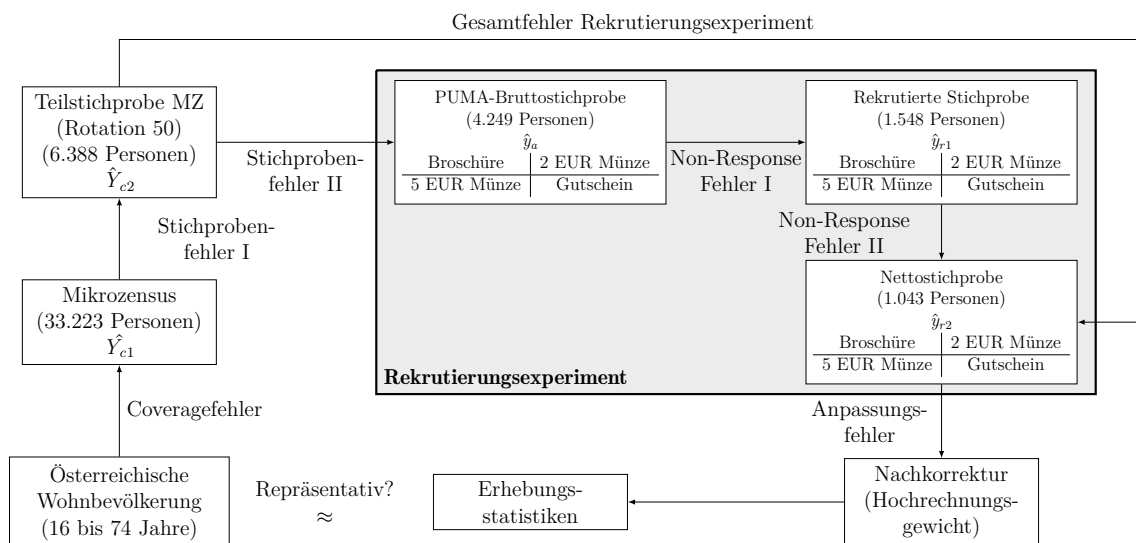


punkt. Ein Vergleich der Incentive-Gruppen nach den Kosten ist rechnerisch aus der notwendigen Stichprobengröße ableitbar. Die statistisch anspruchsvollere und für den Evaluationsbericht relevante Frage umfasst die Genauigkeit und Verlässlichkeit der Daten.

Die Genauigkeit und Verlässlichkeit von Daten lässt sich theoretisch von der klassischen Testtheorie ableiten, wonach jede Messung mit einem Fehler einhergeht. Dieser Fehler kann in den verschiedenen Phasen der Datenerhebung entstehen und in der Literatur wird die systematisch Betrachtung aller Fehlerquellen als Total Survey Error (TSE) bezeichnet (bspw. R. M. Groves et al., 2009; R. M. Groves & Lyberg, 2010; Lyberg & Weisberg, 2016). In der wissenschaftlichen Debatte sind unterschiedliche Differenzierung der Fehlerquellen gebräuchlich. Weisberg (2009) unterscheidet Rekrutierungsprobleme (Stichprobenfehler, Coverage, Unit-Non-Response), Probleme mit der Antwortgenauigkeit (Item-Non-Response, Messfehler aufgrund des Befragten, Messfehler aufgrund des Interviewers) und Probleme bei Befragungsverwaltung (Fehler nach der Erhebung, Fehler durch Erhebungsmodus, Fehler durch Vergleiche). Alternativ unterscheiden R. M. Groves et al. (2009) zwischen Messfehlern und Repräsentationsfehlern. Auf der Seite des Messfehlers differenzieren die Autoren Validität, im Übergang von theoretischem Konstrukt zu Messinstrument; Messfehler, im Übergang von Messinstrument zu Antwort; und Verarbeitungsfehler, im Übergang von Antwort zu verarbeiteten Datensatz. Mit Bezug auf den Repräsentationsfehler ergibt sich ein Coverage-Fehler, im Übergang von Grundgesamtheit zu Auswahlgrundlage; ein Stichprobenfehler, im Übergang von Auswahlgrundlage zur Bruttostichprobe; der Non-Response-Fehler, im Übergang von Brutto- zur Nettostichprobe und ein Anpassungsfehler, mit der Anwendung von Korrekturen an der Nettostichprobe (Vgl. Abbildung 1).

Die Aufgliederung von Weisberg legt bereits nahe, dass Stichprobenziehung, Coverage und Unit-Non-Response, die relevantesten Fehlerquellen für die Evaluation des Rekrutierungsexperiments sind. Auch in der Differenzierung von Groves et al. nach Messfehler und Repräsentationsfehler rückt der Repräsentationsfehler für die vorlie-

Abbildung 2: Repräsentativität über die Stichprobenziehung im Rekrutierungsexperiment<sup>6</sup>



gende Arbeit stärker in den Fokus. Bei beiden Einteilungen darf nicht vernachlässigt werden, dass es sich um theoretisch abgeleitete Differenzierungen handelt. Gerade im Kontext von Incentives ist ein wichtiger Aspekt, ob die Antwortgenauigkeit über Incentive gesteigert werden kann. Dieser Aspekt würde nach Weisbergs Unterscheidung nicht mit der Selektivität korrelieren, allerdings hat der Incentive eben nicht nur Effekte auf die Rekrutierung selbst, sondern auch auf die Motivation allgemein. Hier müsste auch z.B. der Item-Non-Response mitberücksichtigt werden, welcher in der Evaluation aber nicht behandelt wird. Da es in der gegenständlichen PUMA Piloterhebung allerdings nur eine vergleichsweise geringe Zahl an unvollständigen Fragebögen gab, ist von hoher Motivation auszugehen. Eine andere Fehlerquelle würde sich auf die Qualität der tatsächlich gegebenen Antworten beziehen und hängt stark mit dem Fragebogen selbst zusammen. Diese Fehler gehen aber über die Zielsetzung dieses Berichtes hinaus.

## 6 Evaluationskriterien

Aus dem skizzierten Design des Rekrutierungsexperimentes mit Incentives, der Wahl der Webbefragung als Modus der Erhebung, der Stichprobenziehung aus dem Mikrozensus heraus und den vorgestellten Qualitätskriterien für statistische Daten kann eine Liste an Evaluationskriterien abgeleitet werden, welche die Grundlage für die Unterscheidung der verschiedenen Incentive-Gruppen bilden soll. Zur systematischen Einteilung werden die Repräsentationsfehler nach Groves herangezogen, wobei die einzelnen Aspekte durch die Designspezifika ergänzt werden.

<sup>6</sup>Die Bezeichnung der Fehler dient an dieser Stelle einer Verknüpfung des Designs von dem Rekrutierungsexperiment mit dem TSE Schema nach R. M. Groves et al. (2009, , vgl. auch Abbildung 1) und der Benennung der unterschiedlichen Teilaspekte der Fehler in den unterschiedlichen Stichproben, um diese im Anschluss getrennt diskutieren zu können. Als solches reflektiert die Einteilung

Abbildung 2 ist eine an das Rekrutierungsexperiment angepasste Darstellung der Repräsentativitätsproblematik abgeleitet von der Systematisierung durch Groves et al. Es ist an dieser Stelle wichtig darauf hinzuweisen, dass die Bezeichnung als Fehler idealtypisch verwendet wird. Einzig eine Vollerhebung würde Repräsentativitätsprobleme minimieren und jede Stichprobenziehung unabhängig unter welchen Kriterien impliziert die theoretische Möglichkeit eines Fehlers. Der TSE Ansatz betrachtet diesen theoretischen Fehler um potentielle Probleme systematisch zu betrachten. Die Evaluation wird dieser Systematisierung folgen, auch wenn an einzelnen Stellen die Betrachtung über das Rekrutierungsexperiment hinausgeht oder der Fehler nur theoretisch diskutiert werden kann. Der wesentliche Mehrwert des TSE Ansatz liegt genau in diesem ganzheitlichen Ansatz begründet.

Als Grundgesamtheit wurde die österreichische Bevölkerung in Privathaushalten im Alter von 16 bis 74 Jahre definiert. Aus der Eingrenzung des Alters ergibt sich eine Stichprobegröße des Mikrozensus von 33.223 Personen. Diese Stichprobe unterliegt (überwiegend geringen) Coverage-, Stichproben- und Nonresponsefehlern und kann daher die österreichische Wohnbevölkerung nicht vollständig fehlerfrei abbilden (Gumprecht & Oismüller, 2013). Im Kontext des Berichts soll dieser durch die Erhebung des Mikrozensus entstandene Gesamtfehler als Coveragefehler bezeichnen, weil dieser Fehler beim Rekrutierungsexperiment über die Entscheidung für den Mikrozensus entstand und damit nur die fehlende Übereinstimmung mit der österreichische Wohnbevölkerung von Interesse ist. Bei dem Coveragefehler spricht man vom Undercoverage, wenn ein Teil der Zielpopulation nicht berücksichtigt wurde und von Overcoverage, wenn nicht zu untersuchende Personen erhoben wurden. Im Rahmen des Rekrutierungsexperimentes kann kein Overcoverage auftreten. Weiters sollte der Coveragefehler aufgrund der aufwendigen Erhebungsverfahren zur Sicherstellung der Repräsentativität des Mikrozensus sehr gering ausfallen.

Die Rotation 50, das sind die im 2. Quartal 2016 ausgeschiedenen Haushalte, sind die für das Rekrutierungsexperiment relevante Teilstichprobe des Mikrozensus. Zusätzlich wurde die Stichprobe nur auf Personen zwischen 16 und 74 Jahren eingeschränkt. Die 6.388 Personen entsprechen 19,2 % aller Personen des Mikrozensus. Der Stichprobenfehler I bedeutet, dass Schätzer aus dieser Teilstichprobe nicht exakt mit den Schätzern der Gesamtstichprobe übereinstimmen würden. Da die Teilstichproben des Mikrozensus unabhängig voneinander gleich hohe Qualitätsstandards genügen wie der Mikrozensus insgesamt, sollten diese Abweichungen nicht systematisch sein. In Tabelle 45 (Spalte 1: Mik -> Tei) im Anhang finden sich die Ergebnisse einer logistischen Regression, welche dies belegt.

Aus dieser Teilstichprobe des Mikrozensus wurden jedoch nur per CATI direktbefragte Personen zwischen 16 und 74 Jahren, bei denen keine fremdsprachigen Erhebungsunterlagen eingesetzt wurden, aufgenommen. Diese Einschränkungen sind inhaltlich begründbar, aber trotzdem muss geprüft werden, welche Auswirkungen diese Entscheidung hat, denn in diesem Schritt sind 33,5 % der Personen ausgeschieden (Stichprobenfehler II).<sup>7</sup>

---

die Gegebenheiten des Experiments nicht perfekt, soll aber jeweils die inhaltliche Nähe zu der TSE-Terminologie aufzeigen, auch wenn diese nicht im strengen Sinne verstanden werden darf.

<sup>7</sup>Es muss hier darauf hingewiesen werden, dass dieser Fehler auch Non-Response enthält, weil es auch Personen gab, welche nicht erreichbar waren und deshalb nicht als Stichprobeneffekt bezeichnet



Mit der Bruttostichprobe beginnt das eigentliche Rekrutierungsexperiment, wobei die Betrachtung des Coverage- und des Stichprobenfehlers die Annahme prüfen, dass die Bruttostichprobe die Grundgesamtheit hinreichend gut repräsentiert. Mit den bereits beschriebenen Rekrutierungsfragen wurde versucht, die Personen der Bruttostichprobe am Ende der Mikrozensusbefragung zur Teilnahme an der Onlineumfrage zu motivieren. Der durch diese Rekrutierung entstandene Repräsentativitätsfehler soll als Non-Responsefehler I bezeichnet werden, weil er sich von einem zweiten statistisch differenzieren lässt. Die 36,4 % aller Personen in der Bruttostichprobe, die eine Bereitschaft zur Teilnahme zeigten, werden als rekrutierte Stichprobe bezeichnet.

Eine zusätzliche Unterscheidung scheint hier sinnvoll, da die Personen in der rekrutierten Stichprobe einerseits das Incentive erhalten haben, d.h. die Mechanismen des sozialen Austauschs oder alternativer Erklärungsmodelle sollte handlungsrelevant sein, und andererseits durch Erinnerungsschreiben bzw. -E-Mails zusätzlich zur Teilnahme motiviert wurden. Ein weiterer Aspekt im Zusammenhang mit dem als Non-Responsefehler II bezeichneten Selektivitätsproblem ist der Item-Non-Response, der in der Betrachtung aber ausgeklammert wird. Personen mit mehr als 15 fehlenden Antworten wurden als Unit-Non-Response behandelt. Von den zur Befragung letztlich eingeladenen Personen, stehen für 67,4 % nutzbare Daten zur Verfügung.

Die Daten der Nettostichprobe können ähnlich wie die Mikrozensusdaten über Gewichtung an verschiedene Merkmale gebunden werden. Die Gewichtung der Mikrozensusdaten basiert auf einer Anpassung an die Personen in den Bundesländern, Geschlecht, Alter, Staatsbürgerschaft, Erwerbsstatus und Haushaltsgröße. Die für die PUMA-Umfrage berechnete Gewichtung korrigiert entlang Altersgruppen, Bildung, Geschlecht, Erwerbsstatus und Urbanisierungsgrad. Gewichtung wird angewendet, um Repräsentativitätsfehler zu kompensieren, allerdings kann dies auch die Quelle eines Fehlers sein, denn die Varianzen aller Variablen werden künstlich verändert. Die Beurteilung dieser Verzerrung kann über den Vergleich von Mittelwerten des Mikrozensus und der Nettostichprobe erfolgen. Allerdings besteht hier die Problematik, dass das Hochrechnungsgewicht wesentliche sozio-demographische Variablen berücksichtigt und damit der Vergleich entlang dieser Variablen nicht sinnvoll ist. Dementsprechend muss die Prüfung entlang alternativer Variablen stattfinden, welche sowohl für den Mikrozensus als auch die Nettostichprobe zur Verfügung stehen. Der Abbildung 2 folgend wird die Evaluation schrittweise vollzogen. Dabei ermöglicht das Design des Rekrutierungsexperimentes die Quantifizierung von verschiedenen Schätzern zum TSE, welche oft nicht möglich sind. Im Rahmen des Experiments stehen Informationen zum Mikrozensus, zur Teilstichprobe, zur Bruttostichprobe, zur rekrutierten Stichprobe und zur Nettostichprobe zur Verfügung. Dies ermöglicht eine sehr detaillierte Betrachtung und Aufschlüsselung der einzelnen Fehlerkomponenten.

---

werden können. Trotzdem wird hier von Stichprobenfehler gesprochen, weil der Anteil an Personen, welche über die genannten Kriterien ausgeschlossen wurden, deutlich höher ist.

## 6.1 Ausschöpfung

Als grundlegendes Maß zur Evaluierung soll die Ausschöpfungsquote der einzelnen Schritte dienen. Die Ausschöpfungsquote wurde lange Zeit als primäres Qualitätskriterium dokumentiert. Die American Association of Public Opinion Research (AAPOR, 2016) veröffentlicht regelmäßig eine differenzierte Empfehlung für die Definition der unterschiedlichen Gründe für das Ausscheiden bzw. nicht Erreichen von zur Befragung adressierten Personen. Daraus ergeben sich unterschiedliche Berechnungen von Ausschöpfungsquoten, wobei an dieser Stelle die Ausschöpfungsquote allgemein als das Verhältnis von realisierte Ausschöpfung und Ausschöpfungsgrundlage definiert werden soll:

$$\text{Ausschöpfungsquote} = \frac{\text{realisierte Ausschöpfung}}{\text{Ausschöpfungsgrundlage}}. \quad (1)$$

Im Rahmen des Rekrutierungsexperimentes ist diese Definition identisch zu RR6 aus den AAPOR (2016) Empfehlungen, denn durch die Mikrozensusbasis ist sichergestellt, dass der Haushalt bewohnt ist bzw. an der Adresse eine Person anzutreffen ist. Die Ausschöpfungsquote reflektiert aber nicht notwendigerweise den damit einhergehenden Fehler. R. Groves (2006) argumentiert, dass die Qualität einer Umfrage nicht in einer Zahl bemessen werden kann, da der Fehler über jede einzelne Variable variiert. Dementsprechend ist die Ausschöpfungsquote allein kein guter Schätzer für Datenqualität, sondern muss durch die detaillierte Analyse der Ausschöpfungsquote (bspw. Subgruppenvergleiche) und Vergleiche mit externen Schätzern ergänzt werden. Schouten, Cobben und Bethlehem (2009) definieren den Non-Response-Bias mathematisch als:

$$B(\bar{y}) = \frac{\text{cor}(Y, \rho)S(y)S(\rho)}{\bar{\rho}} \quad (2)$$

mit  $\text{cor}(Y, \rho)$  als Korrelation zwischen Zielvariable und Antwortwahrscheinlichkeit,  $S(y)$  als Standardabweichung für die Zielvariable,  $S(\rho)$  als Standardabweichung der Antwortwahrscheinlichkeit und  $\bar{\rho}$  als mittlere Antwortwahrscheinlichkeit. Hierbei lassen sich drei Situation unterscheiden:

- Missing Completely at Random (MCAR):  $\text{cor}(Y, \rho) = 0$  für alle Variablen
- Missing at Random (MAR):  $\text{cor}(Y, \rho) \neq 0$  für andere Variable
- Not Missing at Random (NMAR):  $\text{cor}(Y, \rho) \neq 0$  und  $S(\rho) > 0$  für Zielvariable.

Damit wird deutlich, dass niedrige Ausschöpfungsquoten theoretisch unproblematisch sein können, solange der Unit-Non-Response zufällig ist und nicht mit untersuchungsrelevanten Variablen korreliert. In diesem Fall wird die Repräsentativität der Stichprobe durch die Unit-Non-Response nicht beeinflusst. Deshalb rückte in der Umfrageforschung zunehmend der Non-Response-Bias in den Mittelpunkt der Aufmerksamkeit (R. Groves, 2006). Ist dieser sehr gering, dann sind niedrige Ausschöpfungsquoten

unproblematisch. Der Non-Response-Bias kann deskriptiv bivariat über ausdifferenzierte Ausschöpfungsquoten beschrieben werden, wobei sich hier die aus der Forschung zu Webbefragungen bekannten Attribute anbieten (Erwerbstätigkeit, Bildung, Einkommen, Alter). Außerdem kann über eine binär logistische Regression der Einfluss unterschiedlicher Merkmale multivariat überprüft werden.

## 6.2 R-Indikator

Schouten et al. (2009) entwickelten ausgehend von obigen Überlegungen zu den Ausschöpfungsquoten den sogenannten R-Indikator, um die Repräsentativität einer Stichprobe auch metrisch über die Ausschöpfungsquote hinaus zu analysieren und in einer Kennzahl zu erfassen. Die grundlegende Annahme einer starken Repräsentativität im Sinne des R-Indikators ist, dass jedes Individuum die gleiche Antwortwahrscheinlichkeit<sup>8</sup> aufweist. Dies sollte sich darin ausdrücken, dass die auf Basis des logistischen Regressionsmodells vorausgesagten Antwortwahrscheinlichkeiten eine möglichst geringe Varianz aufweisen. Die Abweichung von diesem Zustand drückt der R-Indikator in einem standardisierten Intervall von 0 (sehr schlecht) bis 1 (sehr gut) aus. Mathematisch wird der globale R-Indikator aus der Standardabweichung der Antwortwahrscheinlichkeit abgeleitet (vgl., de Heij, Schouten & Shlomo, 2010, p. 6):

$$S(\rho) = \sqrt{\frac{1}{N-1} \sum_{i=1}^N d_i (\rho_i - \bar{\rho})^2}. \quad (3)$$

mit gewichteten  $d_i$  als Gewichtungsfaktor<sup>9</sup> der Person  $i$  und  $\rho_i$  als Antwortwahrscheinlichkeiten der Person  $i$ . Damit ist  $S(\rho)$  die gewichtete Standardabweichung der Antwortwahrscheinlichkeit und da  $S(\rho) \leq \frac{1}{2}$  gilt, wird der R-Indikator auf das Intervall von 0 bis 1 standardisiert (vgl., de Heij et al., 2010, p. 6):

$$R(\rho) = 1 - 2S(\rho). \quad (4)$$

Demnach ist die Repräsentativität umso größer, je kleiner die Streuung der Antwortwahrscheinlichkeiten ist. Die Formeln unterstellen, dass die Antwortwahrscheinlichkeiten bekannt sind, da dies aber nicht der Fall ist, wird diese über Hilfsvariablen auf Basis eines binär logistischen Regressionsmodells geschätzt. Der R-Indikator ist demnach ein standardisiertes Streuungsmaß für die Antwortwahrscheinlichkeiten, welches

---

<sup>8</sup>Schouten et al. (2009) interessierte primär der Non-Response und deshalb wird dort auf die Antwortwahrscheinlichkeit verwiesen. Der R-Indikator wird in der Evaluation auf die verschiedenen Auswahlstufen umgelegt, deshalb müsste korrekter von Auswahlwahrscheinlichkeit gesprochen werden. Zugunsten einer verkürzten Terminologie werden Auswahl- und Antwortwahrscheinlichkeit im Rahmen der Diskussion zum R-Indikator synonym verwendet.

<sup>9</sup>Der hier zu verwendende Gewichtungsfaktor berücksichtigt das Stichprobendesign. „The user also needs to define the sample design weights, i.e. the inverse of the sample inclusion probabilities, for all sample units (respondents and non-respondents). For simple random sampling, piinv is equal to 1/pi which is the popsize/samsize defined in step 1. For any other design, the design weight d should be included on the dataset and piinv is equal to d.“(de Heij et al., 2010, p.4)

einen Vergleich über Stichproben hinweg ermöglicht. Die Ergebnisse des R-Indikators hängen natürlich von der Verfügbarkeit von Hilfsvariablen und der Anzahl der Ausprägungen ab. Da die Wahrscheinlichkeiten jeweils für Gruppen geschätzt werden, spielt auch die Besetzungszahl jeder Gruppe eine Rolle. Neben dem globalen R-Indikator können die Ergebnisse des R-Indikators auch partiell ausdifferenziert werden.<sup>10</sup> Dies wird im Rahmen der Evaluation nicht berücksichtigt, denn es werden alle Ergebnisse der logistischen Regressionen entweder im Text oder im Anhang tabellarisch dargestellt und daraus lassen sich die inhaltlichen Ergebnisse ebenfalls ableiten.

Da der R-Indikator eine Schätzgröße ist, wird auch ein Vertrauensintervall bestimmt und zusätzlich werden der maximale Bias und der maximale Kontrast herangezogen, um die Streuung der Antwortwahrscheinlichkeiten abzubilden. Der maximale Bias ist definiert durch (vgl., de Heij et al., 2010, p. 17)<sup>11</sup>:

$$\hat{B}_m(X) = \frac{1 - R(\rho)}{2\bar{\rho}} \quad (5)$$

In dieser Kennzahl, des maximalen absoluten standardisierten Bias einer Variable  $X$ , wird neben der Repräsentativität ( $R(\rho)$ ) auch die durchschnittliche Ausschöpfungsquote ( $\bar{\rho}$ ) über alle Gruppen berücksichtigt unter der Annahme, dass der Non-Response maximal mit den Prädiktoren des Modells der logistischen Regression korreliert. Wenn die Ausschöpfungsquote größer wird, dann verringert sich der maximale Bias. Die Kennzahl ist Null wenn 100 % Repräsentativität erreicht ist und ein kleinerer Wert steht für einen geringeren Bias. Ähnlich verhält es sich mit dem maximalen Kontrast. Dieser kann als die erwartete Differenz zwischen Mittelwert der gültigen Antworten und dem Mittelwert des Non-Response betrachtet werden. Ein Wert von Null würde bedeuten, dass die Antwortwahrscheinlichkeiten identisch über Response und Non-Response sind. Der maximale Kontrast ist definiert als (vgl., de Heij et al., 2010, p. 18):

---

<sup>10</sup>Einerseits können R-Indikatoren für die einzelnen Variablen bestimmt werden, bspw. kann jeweils ein R-Indikator für die einzelnen Items aus der PUMA-Befragung berechnet werden. Andererseits, kann der R-Indikator noch über die unabhängige Variablen und deren Ausprägungen ausdifferenziert werden (see, de Heij et al., 2010, pp. 8-16). Damit kann der Einfluss der einzelnen Variablen oder einzelner Ausprägungen sehr detailliert herausgearbeitet werden. Der Einfluss der einzelnen Variablen lässt sich dabei ungebunden (unconditional partial indicators) oder kontrolliert auf alle anderen Variablen im logistischen Regressionsmodell (conditional partial indicators) bestimmen. Der ungebundene partielle R-Indikator erfasst die Streuung der Antwortwahrscheinlichkeiten zwischen den Kategorien von Variablen. Je stärker die Streuung zwischen den Kategorien ist, desto stärker ist der Einfluss auf die Antwortwahrscheinlichkeit. Der gebundene partielle R-Indikator berechnet den relativen Einfluss kontrolliert auf alle anderen Variablen im Modell.

<sup>11</sup>Die Gleichung 5 impliziert die Annahme, dass die Antwortwahrscheinlichkeit maximal mit den unabhängigen Variablen korreliert - deshalb auch die Bezeichnung als maximaler Bias (Für eine Herleitung siehe Schouten et al., 2009, S. 107). Bei einem perfekten R-Indikator von 1 ist der Bias gleich 0 und falls der R-Indikator nicht perfekt ist, dann ist der Bias zusätzlich über die mittlere Antwortwahrscheinlichkeit ( $\bar{\rho}$ ) spezifiziert. Die Gleichung verdeutlicht auch, dass eine durchschnittliche Antwortwahrscheinlichkeit größer als 0,5 den Zähler positiv zugunsten eines kleineren Bias beeinflusst. Falls  $\bar{\rho} < 0,5$ , dann vergrößert sich der Zähler.

$$\hat{C}_m(X) = \frac{1 - R(\rho)}{2\bar{\rho}(1 - \bar{\rho})} \quad (6)$$

Damit kann der R-Indikator in Ergänzung zur Ausschöpfungsquote und der logistischen Regression auf die Ausschöpfungsquote zur Beurteilung der Repräsentativität herangezogen werden. Allerdings ist der R-Indikator allein nicht hinreichend, sondern es müssen nach den Empfehlungen der Autoren des RISQ-Projekts (de Heij et al., 2010) auch das Vertrauensintervall, der maximale Bias und der maximale Kontrast betrachtet werden, um einen Vergleich von Stichproben zu ermöglichen.

### 6.3 Effektive Stichprobengröße

Der Anpassungsfehler hat im Rahmen des Rekrutierungsexperimentes Relevanz aufgrund der Stichprobenziehung des Mikrozensus. Die Daten werden über die Bundesländer disproportional erhoben zugunsten eines Oversamplings kleiner Bundesländer. Außerdem führt die Befragung aller Haushaltsmitglieder ab 15 Jahre zu Klumpungseffekten. Aufgrund dieser beiden Phänomene ist bei der Betrachtung der Stichprobengröße zu beachten, dass eine reine Zufallsauswahl kleiner ausfallen kann, um Schätzer gleicher Qualität für die Grundgesamtheit zu generieren. Dieser sogenannte Designeffekt<sup>12</sup> kann mathematisch wie folgt definiert werden (Lumley, 2016):

$$D_{eff} = \frac{s_{complex}^2}{s_{random}^2} \quad (7)$$

mit  $s_{complex}^2$  als Varianz einer Variable  $y$  in der komplexen Stichprobe und  $s_{random}^2$  der Varianz der Variable  $y$ , wenn diese aus einer einfachen Zufallsauswahl stammen würde. Aus der Formel ergibt sich auch direkt die Berechnung der effektiven Stichprobengröße ( $n_{eff}$ ) aus der Nettostichprobe ( $n_{netto}$ ):

$$n_{eff} = \frac{n_{netto}}{D_{eff}} \quad (8)$$

Die Voraussetzung zur Berechnung des Designeffekts nach Lumley ist die Möglichkeit, die Varianz unter Berücksichtigung des komplexen Stichprobendesigns zu bestimmen. Außerdem kann der Designeffekt nicht für den gesamten Datensatz bestimmt werden, sondern nur für die einzelnen Variablen.

Beim Fehlen der relevanten Informationen oder zur Schätzung eines gesamten Designeffekts kann die Näherungsformel von Kish (1965) verwendet werden:

$$\hat{n}_{eff} = \frac{(\sum_{i=1}^n w_i)^2}{\sum_{i=1}^n w_i^2} \quad (9)$$

---

<sup>12</sup>Die Schätzung des Designeffekts erfolgt durch das survey Packet in R (Lumley, 2016, 2004).

mit  $w_i$  als Gewicht der Person  $i$ . Diese Vereinfachung kann jedoch Genauigkeitsverluste aufgrund von Klumpeneffekten nicht berücksichtigen.

Andererseits werden allfällige Genauigkeitsgewinne aufgrund einer Korrelation von Gewichten und Erhebungsmerkmalen, wie beispielsweise bei einer stark informativen, gebundenen Hochrechnung nicht berücksichtigt.

## 6.4 Erhebungskosten

Die Kosten werden als abschließendes Evaluationskriterium herangezogen, um basierend auf den effektiven Stichprobengrößen die notwendigen Stichprobengrößen abzuleiten, die dann in den Kosten verglichen werden können. Die notwendige Stichprobengröße wird sowohl unter Berücksichtigung des Designeffekts als auch der Ausschöpfungsquoten in Referenz zu einer Zufallsstichprobe mit einer realisierten Nettostichprobe von 1000 Personen bestimmt. Dementsprechend berechnet sich die notwendige Stichprobe:

$$n_{notw} = \frac{1000 \times D_{eff}}{AQ_{rek} \times AQ_{bru}} \quad (10)$$

mit  $D_{eff}$  als Designeffekt,  $AQ_{rek}$  für die Ausschöpfungsquote der rekrutierten Stichprobe und  $AQ_{bru}$  für die Ausschöpfungsquote der Bruttostichprobe. Für die Incentives werden geschätzte Kosten von 3 EUR für die Broschüre, 5 EUR für die 2 EUR Münze, 8 EUR für die 5 EUR Münze und 11 EUR für den 10 EUR Gutschein angenommen. Die Kosten für die telefonische Rekrutierung werden mit zusätzlich 2 EUR veranschlagt.<sup>13</sup> Der Gutschein ist verhältnismäßig günstig in der Verarbeitung, da hier nur Versandkosten entstehen. Die Münze wird aufwändig per Hand verpackt, was höhere Kosten verursacht. Die Kosten für die Broschüre setzen sich aus Druck-, Verpackungs-, und Versandkosten zusammen. Damit ist die Tabelle 3 als Evaluationsmatrix vollständig.

## 6.5 Variablen für die Untersuchung von Selektivitätseffekten

Die Variablen sind basierend auf einer bivariaten Analyse entlang von Kreuztabellen, welche im Anhang 10.2 dargelegt sind, entsprechend dichotomisiert worden. Dabei wurden die Variablen in Kreuztabellen über die Ausschöpfung von Teilstichprobe Mikrozensus (Rotation 50) auf Brutto- und Nettostichprobe und von Brutto- auf Nettostichprobe untersucht. Basierend auf den Ergebnissen wurden die Variablen so kategorisiert, dass Ausprägungen, die sich nur gering unterscheiden zusammengefasst wurden und die neu gebildeten Kategorien die Unterschiede widerspiegeln. Das Geschlecht aufgrund fehlender Unterschiede von der Analyse ausgeschlossen (Vgl. Tab.

---

<sup>13</sup>Die ungefähre Kostenschätzung basiert auf einer Auskunft der Statistik Austria. Da die jeweiligen Stückkosten auch von der Gesamtmasse abhängig sind und der Anteil der jeweils auf eine Incentivegruppe anfallenden Arbeitszeit nicht exakt bestimmbar war, müssen hier näherungsweise Größenordnungen verwendet werden.

Tabelle 3: Überblick Evaluationskriterien

	Schritt der Stichprobenziehung				
	Mikrozensus	Teilstichprobe (Rotation 50)	Brutto- stichprobe	Rekrutierte Stichprobe	Nettostichprobe
	↓	↓	↓	↓	↕
	Teilstichprobe (Rotation 50)	Brutto- stichprobe	Rekrutierte Stichprobe	Netto- stichprobe	Mikrozensus
Fehler	Stichprobenfehler I+II		Non-Response-Fehler I+II		Anpassungsfehler
Kennzahlen	Ausschöpfungsquote, Logistische Regression		Ausschöpfungsquote, Beschreibung Bias, R-Indikator, Log. Regression		Mittelwertvergleich, Designeffekt
Mikrozensus- variablen	Alter, Einkommen, Familienstand, Haushaltsgröße, Bildung, Berufstätigkeit, Berufliche Stellung, Staatsbürgerschaft, Herkunft				Einkommen, Famili- enstand,
PUMA- Variablen					Vorstellung vom Le- ben

13, Tab. 24 und Tab. 35). Das Alter wurde in drei Altersgruppen von 16 bis 24 Jahre, 25 bis 64 Jahre und 65 bis 74 Jahre umkodiert (Vgl. Tab. 14, Tab. 25 und Tab. 36). Einkommen wurde über die Einkommensquintile erfasst, welche als metrische Variable behandelt wurden. Der Familienstand wurde in verheiratet und in nicht verheiratet dichotomisiert, wobei verheiratet hier auch die eingetragenen Partnerschaften berücksichtigt (Vgl. Tab. 15, Tab. 26 und Tab. 37). Die Haushaltsgröße wurde in kleine Haushalt mit bis zu 4 Personen und große Haushalte mit 5 und mehr Personen unterschieden (Vgl. Tab. 16, Tab. 27 und Tab. 38). Die Bildung wurde in keine Pflichtschule, Pflichtschule, Matura bzw. Lehre, Hochschule und andere Abschlüsse ausdifferenziert. Die Relevanz der Kategorien keine Pflichtschule und andere Abschlüsse haben sich aus der explorativen Analyse der Kreuztabellen abgeleitet (Vgl. Tab. 17, Tab. 28 und Tab. 39). Die Berufstätigkeit wird in berufstätig, nicht berufstätig, in Pension und unbekannt unterschieden. Auch hier leitet sich die Relevanz der Kategorie Unbekannt aus der explorativen bivariaten Analyse ab (Vgl. Tab. 18, Tab. 29 und Tab. 40). Die berufliche Stellung wurde in Arbeiter/in, abhängig Beschäftigte, unabhängig Beschäftigte und Unbekannt/nicht zutreffend eingeteilt (Vgl. Tab. 19, Tab. 30 und Tab. 41). Die Staatsangehörigkeit war bereits dichotom mit österreichisch Staatsbürgerschaft ja oder nein. Und die Herkunft wurde in EU15 inklusive Österreich und nicht EU15 unterschieden (Vgl. Tab. 21, Tab. 32 und Tab. 43). Es wurde Herkunft und Staatsangehörigkeit berücksichtigt, da beide in allen explorativen Analysen der Kreuztabellen signifikante Unterschiede in der Selektivität aufwiesen. Die Variablen aus der PUMA-Befragung wurden in der originalen Kodierung verwendet. Eine Tabelle mit allen Variablen in der originalen und angepassten Kodierungen findet sich im Anhang 10.1. Die berücksichtigten Variablen für die logistischen Regressionen, auch in der Berechnung der R-Indikatoren, sind für alle Modelle gleich, um die Vergleichbarkeit zu gewährleisten. Etwaige Unterschiede in den Tabellen beruhen lediglich darauf,

dass einzelne Ausprägungen nicht mehr besetzt waren.

## 7 Evaluation

Die Struktur der Evaluation soll dem schematischen Verlauf der Abbildung 2 folgen. Ausgehend von der Teilstichprobe MZ (Rotation 50) soll schrittweise nachvollzogen werden, an welchem Übergang welche Verzerrungen oder Selektivitätsphänomene beobachtet werden können, um letztlich über den Vergleich von Nettostichprobe zu Teilstichprobe MZ (Rotation 50) eine Einschätzung der Einschränkungen vorzunehmen und das Rekrutierungsexperiment beurteilen zu können. Der in Abbildung 2 dargestellte Coveragefehler wird nicht behandelt, weil mit keiner Verzerrung zu rechnen ist.

### 7.1 Stichprobenfehler

Der systematische Stichprobenfehler I ist vernachlässigbar klein mit einem Pseudo- $R^2$  von 0,001 (Vgl. Mik -> Tei Modell in Tabelle 45, S. 66) und einem fast perfekten R-Indikator von 0,975. Dieser Fehler wird an dieser Stelle ausgewiesen, um einen Referenzwert für die unterschiedlichen Kenngrößen bereitzustellen, denn im Vergleich von Stichprobenfehler I und II wird bspw. auch die Relevanz des R-Indikators als ergänzendes Evaluationskriterium zur Ausschöpfungsquote deutlich. Die Teilstichprobe MZ (Rotation 50) kann als genauso repräsentativ wie der Mikrozensus insgesamt betrachtet werden und entsprechend drückt sich die minimale Abweichung in einem nahezu perfekten R-Indikator von 0,975 aus, obwohl die Ausschöpfungsquote mit 19,2 % recht gering ist. Beim Übergang von Teilstichprobe auf Bruttostichprobe wurde eine Ausschöpfungsquote von 66,5 % erreicht und trotzdem ist der R-Indikator mit 0,683 deutlich schlechter. Dies liegt wesentlich darin begründet, dass die Auswahl der Bruttostichprobe aus der Teilstichprobe nicht als Zufallsauswahl betrachtet werden kann, sondern bereits deutliche Selektionseffekte auftreten.<sup>14</sup>

Der Stichprobenfehler II<sup>15</sup> hat besondere Relevanz, denn damit ist die Bruttostichprobe und die Ausgangssituation des Rekrutierungsexperiments definiert. In Abbildung 3 ist die Verzerrung über die Resultate einer multivariaten logistischen Regression auf die Auswahl in die Bruttostichprobe aus der Teilstichprobe heraus mit den Odds Ratios und deren Vertrauensintervalle dargestellt. Im Idealfall sollten keine signifikanten Effekte nachweisbar sein, denn dann würde die Auswahl unabhängig von den aufgenommen Variablen sein. Die Ergebnisse zeigen aber eine deutliche Verzerrung bezüglich Alter, wobei junge Menschen bis 24 Jahre eine geringere Chance besitzen in die Bruttostichprobe zu gelangen. Beim Bildungsabschluss sind die Personen mit Hochschulabschluss oder einem anderen unbekanntem Abschluss überrepräsentiert.

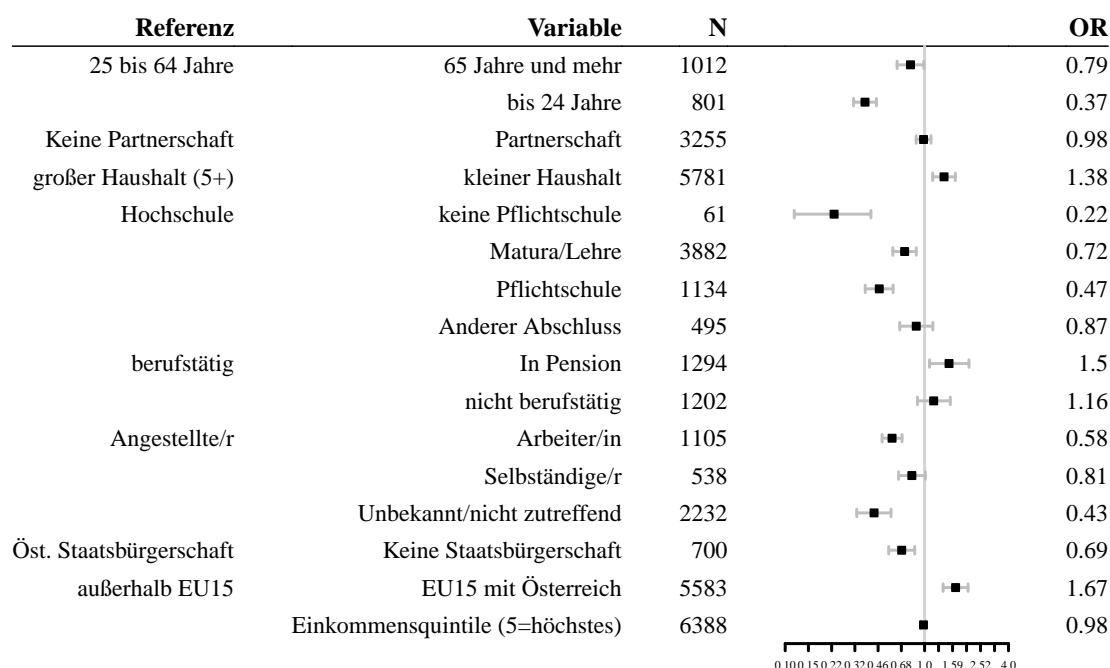
---

<sup>14</sup>Für eine Betrachtung der bivariaten Zusammenhänge der Ausschöpfung sind im Anhang 10.2 die entsprechenden Kreuztabellen aufbereitet.

<sup>15</sup>Dieser Fehler enthält Non-Response, weil es Personen gab, welche telefonisch nicht erreicht wurden. Diese Verzerrung kann streng genommen nicht als Stichprobeneffekt bezeichnet werden. Trotzdem wird hier von Stichprobenfehler gesprochen, weil der Anteil an Personen, welche über die genannten Kriterien (Alter, Sprache, CATI) ausgeschlossen wurden deutlich höher ist.



Abbildung 3: Odds Ratios aus der logistischen Regression auf die Auswahl aus der Teilstichprobe MZ (Rotation 50) in die Bruttostichprobe



Kleine und mittlere Haushalte mit bis zu 4 Personen haben eine erhöhte Chance in der Bruttostichprobe zu sein im Vergleich zu größeren Haushalten. ArbeiterInnen sind genauso unterrepräsentiert, wie Personen welche in einem Land außerhalb der EU15 geboren sind oder keine österreichische Staatsbürgerschaft besitzen. Pensionisten haben eine erhöhte Chance in die Bruttostichprobe aufgenommen zu werden im Vergleich zu berufstätigen Personen. Lediglich Einkommen und Partnerschaft haben keinen signifikanten Effekt auf die Auswahl. Der R-Indikator von 0,683 belegt die starke Verzerrung genauso wie das Pseudo- $R^2$  von 0,081 für die logistische Regression.<sup>16</sup>

## 7.2 Non-Response-Fehler

Der für das Rekrutierungsexperiment relevanteste Fehler ist der Non-Response-Fehler, der in Abbildung 2 in zwei Teile aufgeteilt wurde. Es wurde versucht, die Befragten über das Telefon unter Ankündigung des Incentives zur Teilnahme zu bewegen. Dieser Auswahlsschritt ist durch einen Non-Response charakterisiert, der sich vom Non-Response nach der Zusendung des Incentives unterscheidet. Bei dem Non-Response I handelt es sich zunächst nur um eine Handlungsintention, während der Non-Response II die eigentliche Handlung des hinreichend vollständigen Beantwortens des Fragebogens erfasst. Für die Gesamtbewertung des Non-Response-Fehlers ist letztlich aber auch ein Vergleich von Brutto- und Nettostichprobe nötig. Deshalb werden im folgen-

<sup>16</sup>Die detaillierten Ergebnisse der logistischen Regression mit unstandardisierte Koeffizienten sind im Anhang in der Tabelle 45 auf Seite 66 verfügbar.

Tabelle 4: Ausschöpfungsquote über die Incentivegruppen im Rekrutierungsexperiment

	Gesamt	Broschüre	2 EUR Münze	5 EUR Münze	Gutschein
Bru -> Rek	36,4 %	30,9 %	36,6 %	39,7 %	39,0 %
Rek -> Net	67,4 %	55,2 %	72,1 %	70,2 %	70,4 %
Bru -> Net	24,5 %	17,0 %	26,4 %	27,8 %	27,4 %

den immer drei Übergänge thematisiert: Bruttostichprobe zu rekrutierter Stichprobe (Bru -> Rek), rekrutierte Stichprobe zur Nettostichprobe (Rek -> Net) und Brutto- zur Nettostichprobe (Bru -> Net).

Die Tabelle 4 fasst die Ausschöpfungsquoten über die beiden Selektionsschritte differenziert nach den Incentivegruppen zusammen. Es zeigt sich eine leicht niedrigere Ausschöpfung für die Broschüregruppe im Vergleich zu den Geldincentivegruppen im Übergang zur rekrutierten Stichprobe und ein deutlicher Unterschied in der Ausschöpfung im Übergang zur Nettostichprobe. Die Broschüre wirkt bei der Rekrutierung als auch zur Motivation für die eigentliche Beantwortung schlechter als die Geldincentives, welche sich nur unwesentlich unterscheiden.

Die Ausschöpfungsquoten sind nicht hinreichend als Qualitätskriterium, sondern müssen noch über Verfahren zur Differenzierung der Verzerrungen ergänzt werden. Hier werden zusätzlich die Ergebnisse einer logistischen Regression, bivariate Grafiken und der R-Indikator angeführt. In Tabelle 5 sind potentielle Einflussfaktoren auf die Auswahl für die verschiedenen Schritte in einer logistischen Regression modelliert und signifikante Unterschiede sind erkennbar.

Tabelle 5 stellt die Übergänge von der Bruttostichprobe zur rekrutierter Stichprobe (Non-Response-Fehler I) dar. Bei der Betrachtung des Übergangs von der Brutto- zur Nettostichprobe insgesamt wird deutlich, dass Personen, die in Partnerschaft leben, einen Hochschulabschluss haben, nicht erwerbstätig oder angestellt sind oder innerhalb der EU15 geboren wurden, leichter zu rekrutieren sind. Der Gesamteffekt kann in zwei Arten von Effekten unterteilt werden. Einerseits gibt es Einflussfaktoren die in beiden Auswahlritten einen signifikanten Einfluss haben und manche der Effekte beeinflussen nur einen Auswahlritt. Es gibt mit Bildung, beruflicher Stellung und partiell mit Alter nur drei Faktoren die in beiden Auswahlritten die Selektivität beeinflussen. Bildung ist hierbei der stärkste Effekt, wobei der Unterschied zwischen Personen mit Hochschulabschluss und denen mit höchstens Pflichtschulabschluss besonders deutlich ausfällt. Letztere haben eine geringere Antwortneigung. Der zweite durchgängige Effekt kann bei der Kategorie Arbeiter/in beobachtet werden. Auch hier tritt ein starker Selektionseffekt sowohl im Moment der telefonischen Rekrutierung als auch nach der Zustellung des Incentives auf. Ein interessantes Phänomen ist für die bis 24-Jährigen zu beobachten. Diese Gruppe reagierte scheinbar besser auf die Rekrutierung am Telefon, allerdings nahmen letztlich weniger junge Menschen an der Befragung teil, so dass sich dieser Effekt im Vergleich von Brutto- und Nettostichprobe ausbalancierte.

Bei der zweiten Gruppe an Effekten, welche signifikanten Einfluss auf die Selektivität von Brutto- auf Nettostichprobe besitzen, tritt die Verzerrung entweder bei der tele-

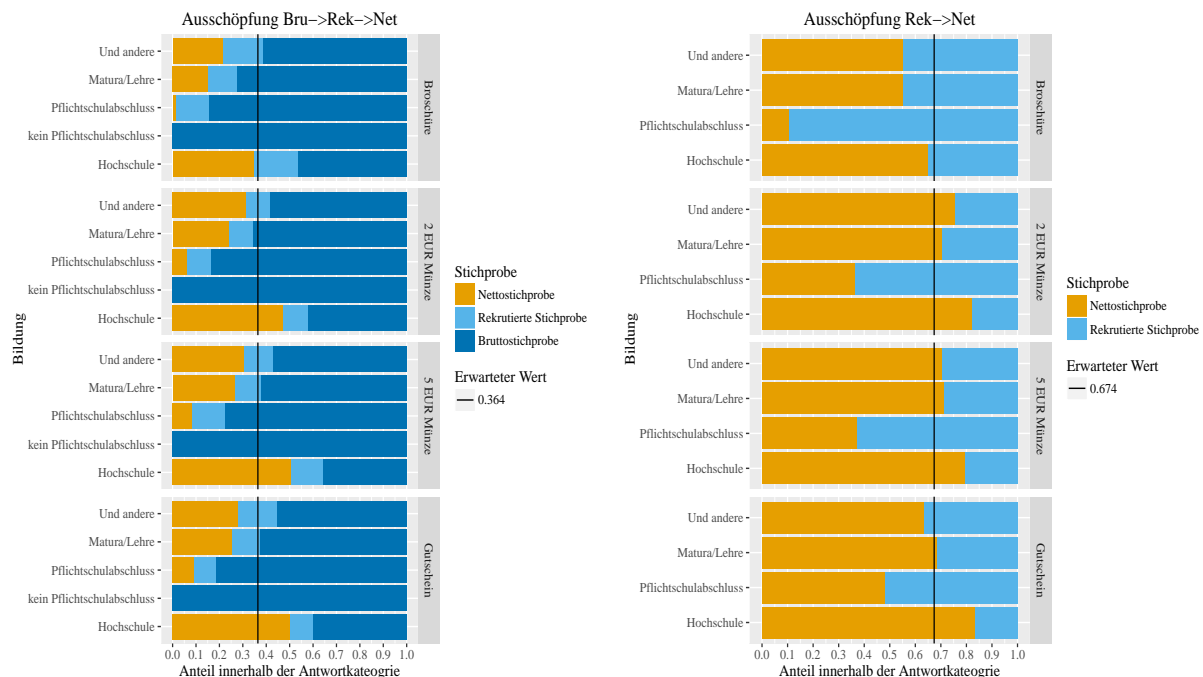
Tabelle 5: Logistische Regressionskoeffizienten der Stichprobenausschöpfung (Von Brutto- zu rekturierter und Nettostichprobe, Standardfehler in Klammern).

PUMA-Variable	Variable in Regression	Chance der Selektion		
		Bru -> Rek	Rek -> Net	Bru -> Net
		<i>logistic</i>	<i>logistic</i>	<i>logistic</i>
	Konstante	-0,567** (0,230)	0,642* (0,377)	-1,275*** (0,267)
Alter (Ref = 25 bis 64 Jahre)	65 Jahre und mehr	-0,287* (0,150)	-0,005 (0,295)	-0,226 (0,175)
	bis 24 Jahre	0,461*** (0,134)	-0,382* (0,213)	0,111 (0,159)
In Partnerschaft (Ref = Nein)	Ja	0,049 (0,072)	0,447*** (0,123)	0,237*** (0,081)
Haushaltsgröße (Ref = Haushalt (5+))	Haushalt (1-4 Personen)	0,187 (0,129)	0,081 (0,224)	0,200 (0,148)
Bildung (Ref = Hochschulabschluss)	Kein Pflichtschulabschluss	-14,024 (224,672)		-13,349 (223,224)
	Matura/Lehre	-0,895*** (0,096)	-0,446*** (0,151)	-0,888*** (0,098)
	Pflichtschulabschluss	-1,485*** (0,148)	-1,547*** (0,262)	-2,067*** (0,202)
	Anderer Abschluss	-0,612*** (0,138)	-0,491** (0,219)	-0,708*** (0,146)
Hauptbeschäftigung (Ref = berufstätig)	In Pension	-0,085 (0,194)	0,058 (0,359)	-0,029 (0,219)
	nicht berufstätig	0,219 (0,150)	0,206 (0,255)	0,294* (0,163)
Berufliche Stellung (Ref = Angestellte/r)	Arbeiter/in	-0,638*** (0,108)	-0,675*** (0,188)	-0,870*** (0,136)
	Selbständige/r	-0,451*** (0,119)	-0,165 (0,206)	-0,432*** (0,131)
	Unbekannte/nicht zutreffend	-0,565*** (0,162)	-0,349 (0,279)	-0,620*** (0,180)
Österreichische Staatsbürgerschaft (Ref = Ja)	Nein	0,163 (0,151)	-0,332 (0,232)	-0,023 (0,173)
Herkunft (Ref = nicht EU15)	EU15 mit Österreich	0,786*** (0,158)	0,379 (0,274)	0,847*** (0,194)
Einkommen	Einkommensquintile (5 = höchstes)	0,042 (0,026)	0,015 (0,044)	0,044 (0,029)
	Pseudo-R <sup>2</sup>	0,074	0,058	0,089
	N	4.249	1.548	4.249
	Log Likelihood	-2.580,822	-921,074	-2.157,775
	Akaike Inf. Crit.	5.195,643	1.874,148	4.349,551

Anmerkung:

\*p<0,1; \*\*p<0,05; \*\*\*p<0,01

Abbildung 4: Balkendiagramm zur Ausschöpfung über Antwortkategorien für Bildung nach Incentivegruppen



fonischen Rekrutierung oder nach der Zustellung des Incentives auf. So ist nach der telefonischen Rekrutierung noch kein Unterschied hinsichtlich der Partnerschaft erkennbar, aber sobald das Incentive eingetroffen ist, zeigen Personen in Partnerschaft eine höhere Teilnahmebereitschaft. Beim Beschäftigungsstatus und der Herkunft tritt der Selektionseffekt bereits bei der Rekrutierung auf, während nach der Zusendung kein Unterschied mehr erkennbar ist.

Auch wenn die logistische Regression die Vorteile der multivariaten Analyse bietet, so ist damit auch der Nachteil verbunden, dass bivariate Zusammenhänge in den Hintergrund treten, sobald die Korrelation zwischen den Kovariaten zu groß ist. In diesem Zusammenhang ist damit zu rechnen, dass Bildung, Beruf, Alter und Einkommen korrelieren und deshalb sind die entsprechenden bivariaten Grafiken für die vier Variablen mit der Ausschöpfungsquote in den Abbildungen 4 bis 7 dargestellt.

Abbildung 4 veranschaulicht den starken Selektionseffekt von Bildung. In der linken Abbildung ist die Ausschöpfungsquote für die Antwortkategorien für die Incentivegruppen von der Brutto- bis zur Nettostichprobe dargestellt. Der Balken für die Nettostichprobe stellt den Anteil des Nettostichprobenumfangs an der Bruttostichprobe dar. Der angrenzende Bereich repräsentiert den Anteil der rekrutierten Stichprobe aus der Bruttostichprobe, welcher nicht in die Nettostichprobe übernommen wurde. Die durchschnittliche Ausschöpfungsquote der Bruttostichprobe in die rekrutierte Stichprobe wird als erwarteter Wert dargestellt. Alle Balken die unter diesem Wert liegen stellen somit unterrepräsentierte Gruppen dar und alle oberhalb des erwarteten Wertes sind in der rekrutierten Stichprobe überrepräsentiert. Um die Betrachtung des Übergangs von rekrutierter Stichprobe auf Nettostichprobe zu verdeutlichen, ist die rechte Abbildung eine Aufskalierung des Übergangs zur Nettostichprobe. Auch

Abbildung 5: Balkendiagramm zur Ausschöpfung über Antwortkategorien für Altersgruppen nach Incentivegruppen

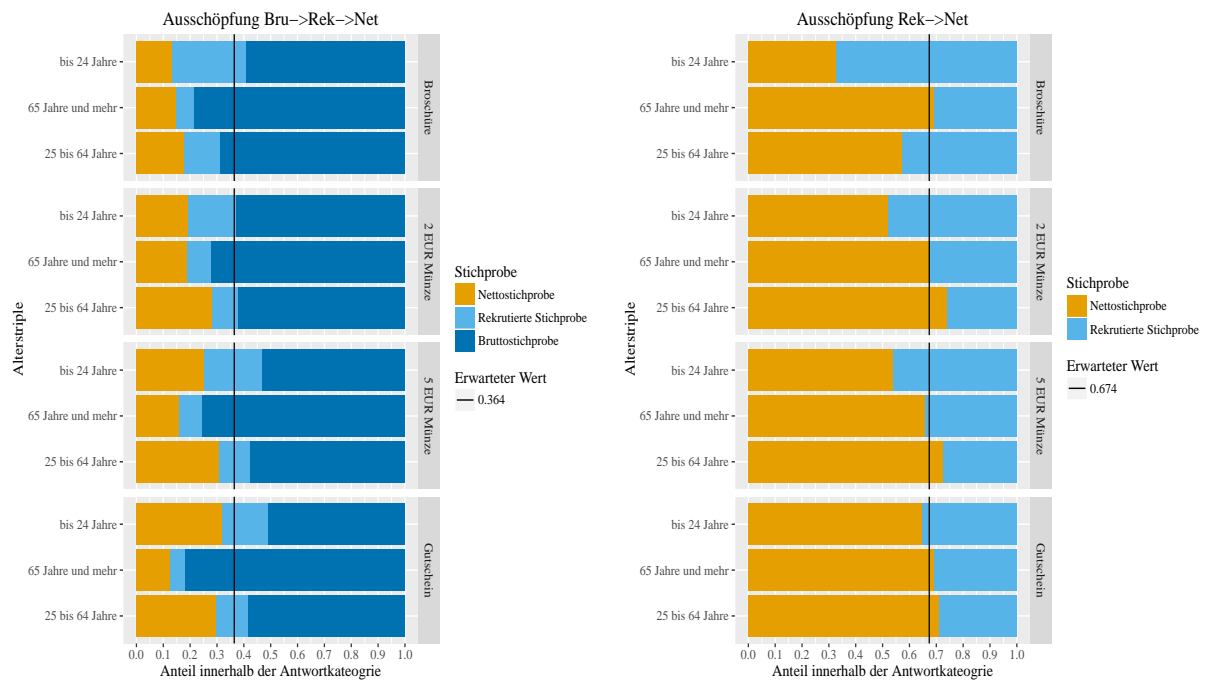


Abbildung 6: Balkendiagramm zur Ausschöpfung über Antwortkategorien für Beruf nach Incentivegruppen

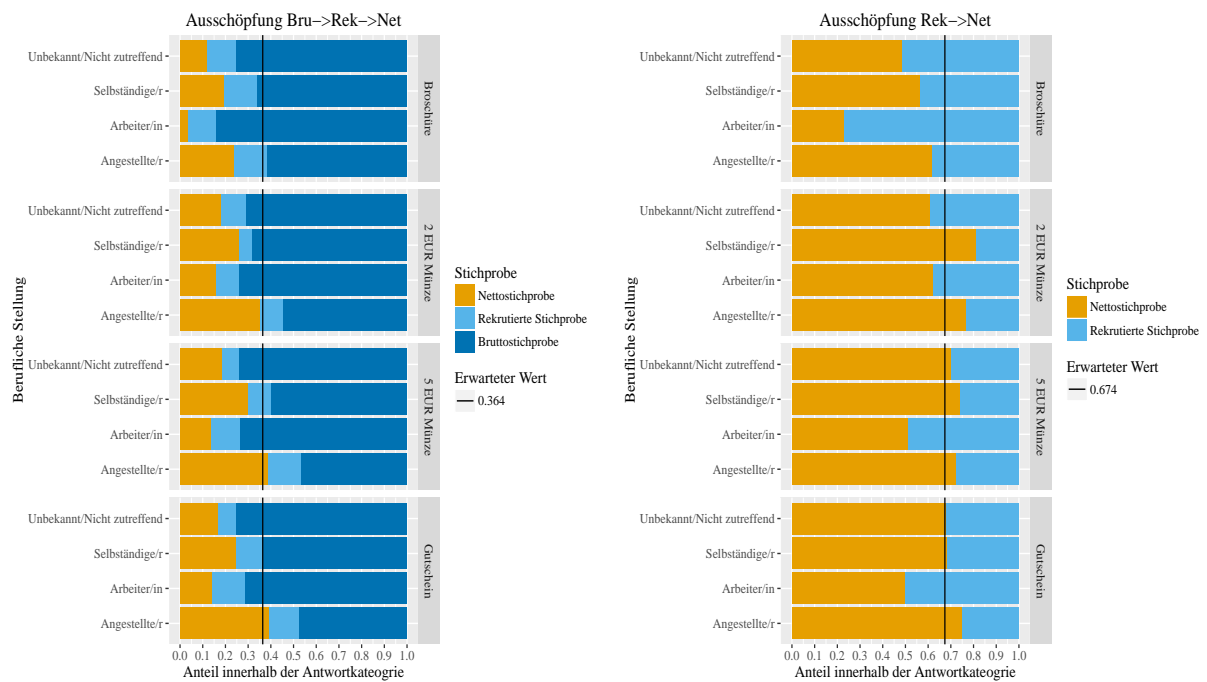
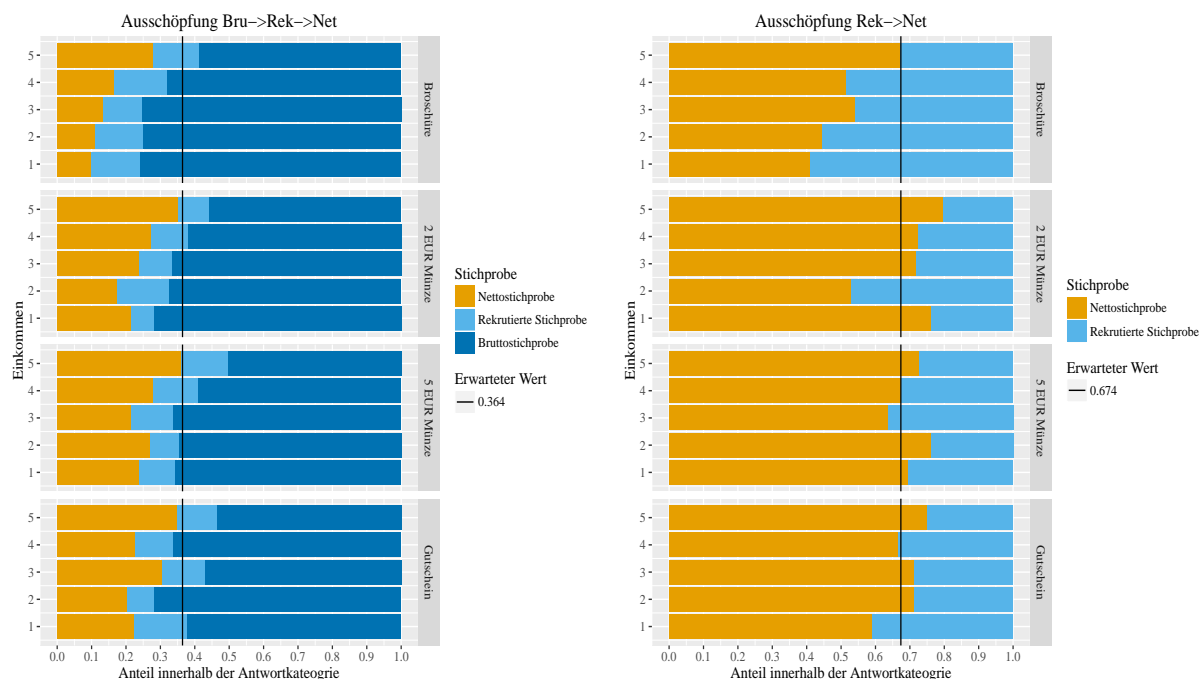


Abbildung 7: Balkendiagramm zur Ausschöpfung über Antwortkategorien für Einkommensquintile nach Incentivegruppen



hier stellt der erwartete Wert die Ausschöpfungsquote der Nettostichprobe aus der rekrutierten Stichprobe ohne Differenzierung nach Incentive oder Variablenausprägung dar. Die Interpretation ist damit recht intuitiv; reicht der am rechten Ende der Abbildung beginnende Balken über die vertikale Linie, dann ist diese Ausprägung in dem Auswahlschritt unterrepräsentiert. Es wird deutlich, dass die Personen mit höchstens Pflichtschulabschluss in beiden Auswahlritten deutlich unterrepräsentiert sind. Gleichzeitig ist auch festzuhalten, dass die Effekte in der logistischen Regression für die anderen Bildungskategorien darin begründet liegen, dass Personen mit Hochschulabschluss deutlich überrepräsentiert sind.

In Abbildung 5 bestätigt sich der Eindruck aus der logistischen Regression, was das besser Ansprechen auf die telefonische Rekrutierung von Personen bis 24 Jahre betrifft, allerdings wird dies durch eine deutlich geringere Beteiligung nach Zusendung der Incentives kompensiert. Bei den Personen die 65 Jahre und älter sind, ist die Situation genau gespiegelt. Diese Gruppe ist am schwierigsten über das Telefon zu rekrutieren, zeigt aber eine nahezu identische Rücklaufquote über alle Incentivegruppen. Insgesamt konnte für die 25 bis 64-Jährigen die beste Ausschöpfungsquote erreicht werden, wobei diese Gruppe sich im ersten Auswahlritt über die Geldincentives besser rekrutieren lässt. Insgesamt ist aber kein signifikanter Alterseffekt im logistischen Regressionsmodell von der Brutto- zur Nettostichprobe nachweisbar. Das interessante Resultat ist hier, dass die Ausschöpfung am stärksten für die Gutscheingruppe variiert, weil die älteren Personen kaum darauf reagieren, während die beiden jüngeren deutlich höhere Ausschöpfungsquoten in der Nettostichprobe zeigen.

Betrachtet man die Ausschöpfungsquoten über die berufliche Stellung (Vgl. Abbildung 6), so kann festgehalten werden, dass Angestellte deutlich leichter zu rekrutieren

Tabelle 6: R-Indikator für die Incentivegruppen über den Verlauf des Rekrutierungsexperimentes

	Gesamt	Broschüre	2 EUR Münze	5 EUR Münze	Gutschein
Bru -> Rek	0,715	0,726	0,746	0,694	0,689
Rek -> Net	0,745	0,698	0,732	0,738	0,742
Bru -> Net	0,730	0,782	0,733	0,703	0,692

sind. Der Unterschied zwischen den Antwortkategorien resultiert hier - mit Ausnahme für Arbeiter/in - auf der telefonischen Rekrutierung. In der zweiten Stufe nach der Zusendung der Incentives, zeigen alle Gruppen eine niedrige Ausschöpfung für die Broschüre, aber vergleichbare Quoten für die Geldincentives. ArbeiterInnen haben bereits bei der telefonischen Rekrutierung deutlich geringere Ausschöpfungsquoten, die für die Broschüre nochmals geringer ausfallen als für die Geldincentives.

Bei der Betrachtung des Zusammenhangs von Einkommen und Ausschöpfungsquote (Vgl. Abbildung 7) wird deutlich, dass bei der Rekrutierung über die Broschüre und den 2 EUR Münze Incentive ein Zusammenhang erkennbar wird: Höhere Einkommensgruppen sind leichter zu rekrutieren und zeigen auch eine besser Teilnahme nach der Zusendung der Incentives. Für die Einkommensquintile eins bis drei lassen sich leicht positive Effekte für höhere Incentives erkennen, allerdings sind die Unterschiede im Vergleich zur 2 EUR Münze nur geringfügig bzw. vereinzelt relevant.

Zum Abschluss der Betrachtung des Non-Response-Fehlers soll noch der R-Indikator herangezogen werden, um die Gesamtverzerrung in einer Kennzahl zusammenzufassen. Tabelle 6 stellt den R-Indikator für die Incentivegruppen über die beiden Auswahlstufen und im Vergleich von Brutto- zu Nettostichprobe dar. Ein R-Indikator von 1 würde eine perfekt repräsentative Stichprobe mit keinerlei Verzerrung bezüglich der Kovariaten im Regressionsmodell charakterisieren. Im Übergang von Bruttostichprobe zur rekrutierten Stichprobe zeigen die Incentivegruppen mit den Broschüren und der 2 EUR Münze eine höhere Repräsentativität als die anderen beiden Geldincentivegruppen. Im Übergang zur Nettostichprobe zeigen sich keine Unterschiede hinsichtlich der Geldincentivegruppen, aber die Broschürengruppe zeigt einen niedrigeren Wert. Etwas überraschend ist, dass die Broschürengruppe den höchsten R-Indikator für den Übergang von Brutto- zu Nettostichprobe hat. Dies lässt sich auch durch gegenläufige Selektionseffekte in den beiden Auswahlritten erklären, verdeutlicht aber auch die Relevanz der Vertrauensintervalle, des maximalen Bias und des maximalen Kontrast, um die Verzerrung genauer einschätzen zu können. Die Vertrauensintervalle zeigen, dass die geschätzte Repräsentativität zwischen den Gruppen sich nicht signifikant unterscheidet. Es ist daher wesentlich neben der Repräsentativität gleichzeitig die durchschnittliche Ausschöpfung in Rechnung zu stellen. Dies ermöglichen die beiden zusätzlichen Indikatoren für den maximalen Bias bzw. Kontrast.

Tabelle 7 fasst nochmals die Ausschöpfungsquoten und die differenzierte Betrachtung des R-Indikators zusammen. Die Geldincentives zeigen deutlich höhere Ausschöpfungsquoten als die Broschüre, während die Unterschiede zwischen den Geldincentives nur marginal sind. Bei der Betrachtung der Vertrauensintervalle wird deutlich, dass sich die R-Indikatoren über die Incentivegruppen nicht signifikant voneinander

Tabelle 7: R-Indikator und Pseudo- $R^2$  der logistischen Regression für Ausschöpfung aus Brutto- zu Nettostichprobe nach Incentivegruppe

	n	Ausschöpfung	R-Indikator	Vertrauensintervall	Maximaler Bias	Maximaler Kontrast	Pseudo- $R^2$
Broschüre	1.115	17,0 %	0,782	(0,733-0,830)	0,609	0,742	0,112
2 EUR Münze	1.046	26,4 %	0,733	(0,680-0,785)	0,473	0,388	0,089
5 EUR Münze	1.056	27,8 %	0,703	(0,650-0,755)	0,499	0,409	0,098
Gutschein	1.032	27,4 %	0,692	(0,638-0,746)	0,530	0,435	0,103

unterscheiden lassen. Außerdem zeigt sich im maximalen Bias und maximalen Kontrast, dass der potentielle Fehler für die Broschürengruppe am höchsten ist, wenn man neben der Repräsentativität auch die Gesamtausschöpfung berücksichtigt. Unter Berücksichtigung von Ausschöpfungsquote und R-Indikatoren zeigt die Gruppe mit der 2 EUR Münze das beste Ergebnis.

### 7.3 Gesamtfehler über das Rekrutierungsexperiment

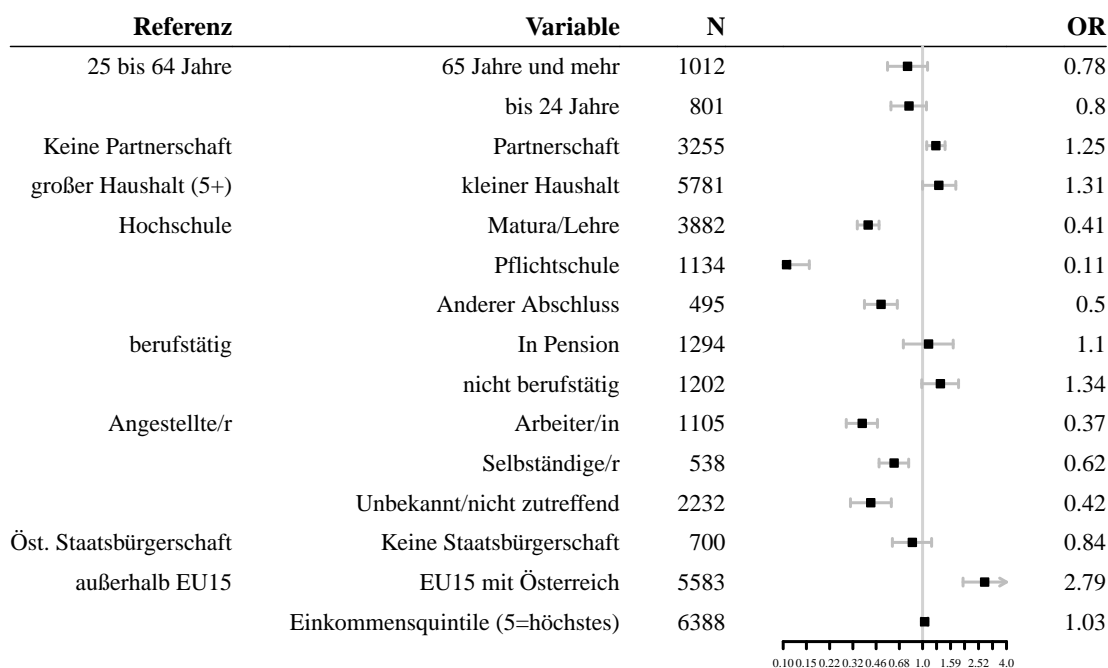
Zur Betrachtung des Gesamtfehlers soll hier die Teilstichprobe des Mikrozensus mit der Nettostichprobe verglichen werden, um die Selektivität der Nettostichprobe im Vergleich zu einer repräsentativen Stichprobe darzustellen. Damit wird ein Vergleich zum Stichprobenfehler möglich und außerdem der gesamte Fehler dargestellt, welcher von der Selektion in die Bruttostichprobe und über das Rekrutierungsexperiment selbst entstanden ist.

Aus Abbildung 8 wird ersichtlich, dass in der Nettostichprobe eine Selektivität hinsichtlich der Bildung, zugunsten der HochschulabsolventInnen und zu Ungunsten der Personen mit Pflichtschulabschluss, eines erhöhten Vorkommens von Angestellten, in Partnerschaft lebenden Personen und Personen aus der EU15 auftritt. Alter, Haushaltsgröße, Berufstätigkeit, Staatsbürgerschaft und Einkommen können vernachlässigt werden.

Im Vergleich zu Abbildung 3 nivellieren sich einige Unterschiede wie Alter, Berufstätigkeit oder Haushaltsgröße aus. Gleichzeitig verschärfen sich die Selektivitätsprobleme mit Bezug auf PflichtschulabsolventInnen oder Personen außerhalb der EU15. Dass die Selektivität durch das Rekrutierungsexperiment insgesamt nicht verbessert wurde, zeigt auch ein Vergleich des Pseudo- $R^2$  und des R-Indikators für beide Auswahlstapen. Der Übergang von Teilstichprobe zu Bruttostichprobe kann durch das Modell zu 8,1 % erklärt und durch einen R-Indikator von 0,683 (VI: 0,66-0,706;  $B_m=0,239$ ;  $C_m=0,709$ ) beschrieben werden. In der Betrachtung der Selektivität der Nettostichprobe aus der Teilstichprobe heraus ergibt sich ein Pseudo- $R^2$  von 11,5 % und ein R-Indikator von 0,757 (VI: 0,736-0,777;  $B_m=0,704$ ;  $C_m=0,851$ ). Nach dem R-Indikator allein wäre die Nettostichprobe sogar repräsentativer als die Bruttostichprobe, allerdings wird beim Vergleich des maximalen Bias und des maximalen Kontrasts deutlich, dass die Nettostichprobe nicht besser ist. Das Pseudo- $R^2$  für die Ausschöpfung der Teilstichprobe Mikrozensus (Rotation 50) in die Nettostichprobe zeigt an, dass ein höherer Anteil der Varianz durch das Regressionsmodell bestimmt, was widersprüchlich zum Ergebnis des R-Indikators erscheint. An dieser Stelle wird deutlich, dass der R-Indikator ein Streuungsmaß für die Antwortwahrscheinlichkeiten ist, während



Abbildung 8: Odds Ratios aus der logistischen Regression auf die Auswahl aus der Teilstichprobe in die Nettostichprobe



das Pseudo- $R^2$  angibt, wie gut diese erklärt wird. Dementsprechend ergänzt der R-Indikator die Analyse nur und muss immer im Kontext von maximalem Bias und maximalen Kontrast interpretiert werden (vgl., de Heij et al., 2010).

## 7.4 Anpassungsfehler

Bis zu diesem Zeitpunkt sind die Daten ohne Gewichtung verwendet worden. Allerdings stellt die Statistik Austria sowohl für den Mikrozensus als auch für die PUMA-Stichprobe einen angepassten Gewichtungsfaktor zur Verfügung. Das Mikrozensusgewicht extrapoliert die Stichprobe auf die Gesamtbevölkerung Österreichs und wurde in einer iterative proportionalen Anpassung an die Variablen Einwohner pro Bundesland, Geschlecht, Alter, Staatsbürgerschaft, Nationalität, Erwerbstätigkeit und Haushaltsgröße gebunden (Meraner, Gumprecht & Kowarik, 2016). Die Ergebnisse des Mikrozensus entsprechen auf diesen Variablen also den wahren Werten der Bevölkerung in Österreich. Das angepasste Hochrechnungsgewicht für die PUMA-Umfrage steht nur für die Nettostichprobe zur Verfügung, und nimmt eine Bindung nach Alter, Bildungsstand, Geschlecht, Erwerbsstatus und Urbanisierungsgrad vor. Für den Mikrozensus bildet die Wohnbevölkerung die Grundgesamtheit, während für die PUMA-Befragung nur die Personen zwischen 16 und 74 Jahren berücksichtigt wurden. Meraner et al. (2016) diskutieren die Verbesserungen in der aktuellen Gewichtung des Mikrozensus und deshalb sollten auch die Abweichungen für die PUMA-Stichprobe marginal sein.<sup>17</sup>

<sup>17</sup>Meraner et al. (2016) führen keine Testung der PUMA-Stichprobe vor, aber aufgrund der vergleichbaren Anwendung der Gewichtung für Mikrozensus und PUMA-Stichprobe sind ähnliche

Tabelle 8: Gewichtete Mittelwerte, Standardfehler und Designeffekte nach Lumley einzelner Variablen aus der PUMA Befragung

	Variable	Mittelwert	Standardfehler	Designeffekt
	Allgemeine Gesundheit	1,94	0,03	1,28
	Vorstellung vom Leben	8,18	0,08	1,37
Vergleich mat. Bedingung heute vs. 10 J.		3,38	0,04	1,53
	Gefühltes Alter	40,31	0,50	1,47

Der Designeffekt über die Variablen aus der PUMA-Umfrage variiert von 1,28 bis 1,53 (Vgl. Tabelle 8). Die Schätzung des Designeffekts nach Kish würde geringere Werte aufweisen, allerdings wird bei diesen Kish-Schätzern die Klumpung nicht berücksichtigt und bei vorhandener Klumpung wird die effektive Stichprobe überschätzt. Insgesamt lässt sich festhalten, dass das Stichprobendesign des Mikrozensus die Nettostichprobe im Vergleich zu einer reinen Zufallsstichprobe nochmals verringert. Der Designeffekt variiert auch über die Incentivegruppen, weil diese unterschiedlich zusammengesetzt sind und beispielsweise selektive Ausfälle durch die gebundene Hochrechnung in unterschiedlichem Ausmaß kompensiert werden müssen. Die Gewichtung in den Incentivegruppen ist für Vergleiche nicht ideal, weil die Hochrechnung auf die gesamte Nettostichprobe erfolgte idealerweise aber auf jede Gruppe erfolgen müsste. Deshalb wird im folgenden Vergleich der Kosten für alle Incentivegruppen ein konstanter Designeffekt von 1,37 angenommen (Vgl. Variable Vorstellung vom Leben in Tabelle 8).

## 7.5 Kosten

Auf der Basis der effektiven Stichprobengröße bzw. den Designeffekten lassen sich Überlegungen zu den Kosten für die unterschiedlichen Incentivegruppen anstellen. Die Tabelle 9 stellt die notwendige Stichprobengröße für die Bruttostichprobe und die rekrutierte Stichprobe dar, um eine Nettostichprobengröße von 1000 Personen zu realisieren, wenn die Ausschöpfungsquoten und der Designeffekt der jeweiligen Incentivegruppe berücksichtigt wird. Ausgehend von der über den Designeffekt angepassten theoretischen Nettostichprobengröße wurden dann die im Rekrutierungsexperiment

Ergebnisse plausibel. Eine Prüfung des gewichteten Mittelwerts für Alter ergab nur eine minimale Abweichung 0,3 Jahren zwischen Mikrozensus und Nettostichprobe.

Tabelle 9: Notwendige Bruttostichprobe bzw. rekrutierte Stichprobe nach Incentivegruppe unter Berücksichtigung der Ausschöpfungsquoten um eine effektive Nettostichprobe von 1000 zu erreichen.

	Effektive Stichprobe	Design-effekt	Notw. Stichprobe mit Designeffekt	Ausschöpfung Rek. Stichprobe	Notw. Rekrutierte Stichprobe	Ausschöpfung Bruttostichprobe	Notw. Bruttostichprobe
Broschüre	1.000	1,37	1.372	55,2 %	2.484	30,9 %	8.050
2 EUR Münze	1.000	1,37	1.372	72,1 %	1.903	36,6 %	5.199
5 EUR Münze	1.000	1,37	1.372	70,2 %	1.955	39,7 %	4.927
Gutschein	1.000	1,37	1.372	70,4 %	1.949	39,0 %	5.002

Tabelle 10: Kostenschätzung basierend auf der ermittelten notwendigen Bruttostichprobe

	Notw. Bruttostichprobe	Telefonkosten (2 EUR pro Anruf)	Stückkosten Incentive	Rekrutierte Stichprobe	Incentivekosten	Gesamtkosten
Broschüre	8.050	16.100 EUR	3 EUR	2.484	7.451 EUR	23.550 EUR
2 EUR Münze	5.199	10.397 EUR	5 EUR	1.903	9.517 EUR	19.915 EUR
5 EUR Münze	4.927	9.854 EUR	8 EUR	1.955	15.639 EUR	25.493 EUR
Gutschein	5.002	10.004 EUR	11 EUR	1.949	21.434 EUR	31.438 EUR

aufgetretenen Ausschöpfungsquoten zur rekrutierten Stichprobe und dann zu Bruttostichprobe hochgerechnet. Für die Kosten wurden einerseits Kosten für die telefonische Rekrutierung von 2 EUR pro Anruf auf die notwendige Bruttostichprobe<sup>18</sup> umgelegt. Andererseits, wurden die Kosten für die Incentives nur für die notwendigen rekrutierten Stichprobengrößen, berechnet auf Basis der Ausschöpfungsquoten des Experiments, mit einbezogen. Tabelle 10 gliedert die Kosten nach den Incentivegruppen auf.

Die Gesamtkosten für die 2 EUR Münze sind am geringsten, was sich im Vergleich zur 5 EUR Münze und dem Gutschein aus den niedrigeren Stückkosten des Incentives erklärt, wobei die Ausschöpfung hier keine wesentliche Rolle spielt. Die Kosten für die Broschüre sind nach der 2 EUR Münze am niedrigsten, d.h. auch die schlechte Ausschöpfungsquoten sind nicht so stark, dass die zusätzlichen Telefonkosten die 5 EUR Münze oder den Gutschein kostengünstiger werden lassen. Allerdings ist anzumerken, dass in die Berechnung nur Teilaspekte der Ausführungen zur Repräsentativität der Stichprobe mit einfließen.

Damit kann mit Blick auf die Kosten festgehalten werden, dass die 2 EUR Münze das günstigste Design ist, um eine effektive Nettostichprobe von 1.000 Personen aus der Mikrozensusstichprobe zu erreichen. Die Broschüre ist etwas mehr als ca. 3.600 EUR teurer, wobei 5.700 EUR Mehrkosten, im Vergleich zur 2 EUR Münze, für die zusätzlichen Telefonanrufe aufgrund der deutlich größeren notwendigen Bruttostichprobe entstehen. Die niedrigeren Incentivekosten für die Broschüre kompensieren diese Kosten nur teilweise. Die deutlichen Unterschied zur 5 EUR Münze und zum Gutschein entstehen durch die höheren Kosten der Incentive.

## 8 Schlussfolgerungen und Empfehlungen

Zunächst zwei theoretische Vorüberlegungen, welche zur Einordnung der Ergebnisse notwendig erscheinen. Die Personen wurden direkt aus der Mikrozensusbefragung heraus rekrutiert. Daraus ergibt sich einerseits das Problem, dass die inhaltliche Trennung beider Befragungen nicht zwangsläufig von allen Befragten realisiert wurde. Es ist also möglich, dass eine höhere Teilnahme daraus resultiert, dass die Befragten immer noch glauben am Mikrozensus teilzunehmen, an welchem die Teilnahme verpflichtend ist. Andererseits ist in der Literatur gut dokumentiert, dass Befragte bei Panelstudien eine Befragungsermüdung aufweisen und damit die Bereitschaft zur Teil-

<sup>18</sup>Aufgrund der konservativeren Schätzung nach Lumley sind die entsprechenden Stichprobengrößen deutlich größer als die Schätzungen basierend auf dem Designeffekt nach Kish wären. Der Designeffekt nach Kish würde 1,27 betragen.

nahme sinkt (Porter, Whitcomb & Weitzer, 2004). Diese beiden Effekte können für die unterschiedlichen Bevölkerungsgruppen jeweils auf verschiedene Art wirken und lassen sich empirisch nicht testen.

Die Ergebnisse der Evaluation zeigen ein durchaus gemischtes Bild, denn eine eindeutig beste Variante kann nur mit Einschränkungen bestimmt werden. Grundsätzlich ist die 2 EUR Münze unter Abwägung von Qualität und Kosten die beste Variante. Danach wird die Situation schon schwieriger, denn die Broschüreggruppe hat zwar die geringste Ausschöpfungsquote und die Verzerrungen in der Repräsentativität sind ebenfalls am stärksten, allerdings ist die Variante günstiger als die 5 EUR Münze und die Gutscheinvvariante. Der geringe Kostenunterschied von 2.000 EUR zwischen Broschüreggruppe und 5 EUR Münzengruppe wiegt die Qualitätsverluste noch nicht auf. Die 5 EUR Münze und die Gutscheinvvariante sind gleichzeitig aber nicht besser in der Qualität der Stichprobe als die Variante mit der 2 EUR Münze. Deshalb kann die 5 EUR Münze als unwesentlich teurer aber qualitativ bessere Variante im Vergleich zur Broschüreggruppe als zweitbeste Incentivegruppe insgesamt betrachtet werden. Die Broschüreggruppe ist qualitativ schlechter als die Gutscheingruppe dafür aber deutlich günstiger in den Kosten. Eine Differenzierung in Rangplatz drei und vier würde entweder den Kosten oder der Qualität Vorrang geben.

Die Geldincentives erreichen eine höhere Ausschöpfungsquote und Repräsentativität als die Sachincentivegruppe. Dabei sind die Unterschiede der Geldincentives mit Bezug auf die Ausschöpfungsquote marginal allerdings unterscheiden sich die Geldincentives teilweise dahingehend, wie bestimmte Subpopulationen darauf reagieren. Die PflichtschulabsolventInnen reagierten positiv auf höhere Geldincentives im Vergleich zu niedrigeren, während bei ArbeiterInnen die Nettostichprobengröße mit steigendem Geldincentive sank.

Im Allgemeinen verringern die Geldincentives die Selektivität im Besonderen für Bildung und Einkommen. Dies ist ein besonders positiver Aspekt des Rekrutierungsexperimentes, denn gerade für Bildung war der Selektivitätseffekt besonders stark. Trotzdem darf der Effekt der Geldincentives nicht überschätzt werden, denn die Verzerrung der Stichprobe über die Bildung bleibt auch nach der leichten Verbesserung durch die Incentives der stärkste Faktor von Selektivität.

Ein wichtiges Resultat der Analyse ist, dass die Bruttostichprobe bereits einen Selektivitätsbias aufwies, so dass die Ausgangsbedingungen für das Rekrutierungsexperiment keine perfekt repräsentative Stichprobe war. Die Entscheidung für die Ausschlusskriterien sind gut nachvollziehbar, allerdings ist es sehr empfehlenswert für eine zukünftige Stichprobenziehung aus einer Teilstichprobe des Mikrozensus, die Selektivität direkt zu prüfen und die Bruttostichprobe entsprechend anzupassen. So lässt sich nicht bestimmen, ob diese Verzerrung nicht auch das gesamte Rekrutierungsexperiment nachhaltig verfälscht hat.

Aus der Analyse lassen sich auch einige Empfehlungen ableiten. Der angesprochene Bias in der Bruttostichprobe ist unbedingt zu vermeiden, denn er basiert auf nachvollziehbaren pragmatischen Gründen, kann aber auch korrigiert werden. Hier sollte Sorge getragen werden, dass eine entsprechende Prüfung vor der telefonischen Rekrutierung stattfindet, damit die Qualität der Stichprobe nicht schon zu Beginn negativ beeinflusst wird.

Das Erhebungsdesign des Mikrozensus mit seiner Klumpung der Haushalte und der

disproportionalen räumlichen Allokation über die Bundesländer verringert die effektive Stichprobengröße teils deutlich. Dies ist besonders im Vergleich zu einer Zufallsstichprobe ein großer Nachteil, denn die Kosten für die notwendige größere Stichprobe werden durch die Telefon- und Incentivekosten spürbar erhöht. Deshalb ist es vom Standpunkt der Erhebungskosten für eine Webbefragung grundsätzlich empfehlenswert, Klumpungs- und Allokationseffekt möglichst gering zu halten. Die um ein Viertel bis zu ein Drittel geringere effektive Stichprobengröße bedeutete zusätzliche Kosten bei der Erhebung sowohl für die AuftraggeberInnen aber auch für die Befragten. Allerdings hat der Mikrozensus den Vorteil, dass die Befragten auch erreichbar sind und dass wenigstens der Rekrutierungsversuch unternommen werden kann. Auch für eine Zufallsstichprobe wäre eine größere Stichprobe ratsam, denn das Rekrutierungsexperiment hat gezeigt, dass einzelne Subpopulationen eine sehr geringe Teilnahmebereitschaft zeigen. Soll für diese Personengruppen hinreichend Informationen zur Analyse bereitstehen, sollte sich die Stichprobengröße am Anteil der Gruppe in der Bevölkerung hochgerechnet über die Ausschöpfungsquoten orientieren.<sup>19</sup>

Eine dritte Empfehlung resultiert aus den Betrachtungen zur Wirksamkeit der Incentives. Auch wenn grundsätzlich das 2 EUR Münzen Incentive als beste Lösung erscheint, so können die Qualität und die Kosten noch durch differenzierte Incentives optimiert werden. Die leicht zu rekrutierenden Gruppen könnten entweder einen geringeren Geldincentive erhalten oder gar die Broschüre, während die schwer zu rekrutierenden Gruppen mit höheren Incentives versucht werden zu motivieren. Die Anwendung von differenzierten Incentives ist in der Literatur durchaus umstritten, denn die differenzierten Incentives widersprechen dem Gleichbehandlungsgrundsatz und können als unfair betrachtet werden (Ernst Stähli & Joye, 2016). Auch wenn diese Variante vom Standpunkt der Gesamtbefragungskosten und damit für die öffentliche Hand als kosteneffizient betrachtet werden kann, so könnte das Instrument in der öffentlichen Debatte wegen der ungleichen Behandlung der TeilnehmerInnen in die Kritik geraten. Da die Gruppen, welche für einen erhöhten Incentive in Frage kommen, aber aus den einkommensschwächeren Gruppen der Gesellschaft kommen und die Gutscheinincentives mit einem Wert von 10 EUR noch überschaubar sind, sollte dieser Effekt mit einer entsprechenden Öffentlichkeitsarbeit kompensierbar sein. Insgesamt kann festgehalten werden, dass die Rekrutierung über den Mikrozensus zu einer Webbefragung mit Hilfe von Incentives eine gute Möglichkeit darstellt die Stichprobenziehung zu kontrollieren, was normalerweise eines der größten Probleme von Webbefragungen darstellt. Auch für die Evaluation standen sehr detaillierte Informationen für die Befragten zur Verfügung und damit war eine genauere Betrachtung von potentiellen Fehlerquellen möglich.

---

<sup>19</sup>Bspw. haben die Personen mit Pflichtschulabschluss einen Ausschöpfungsquote in der Broschü-  
rengruppe von ca. 10 % gehabt. und in der Gutscheingruppe von ca. 40 %. Wenn also in der Netto-  
stichprobe 100 von 1.000 Personen Pflichtschulabschluss haben sollen, dann wäre mit der Broschüre  
eine Bruttostichprobe von 10.000 und mit dem Gutscheinincentive eine Bruttostichprobe von 2.500  
anzustreben.

## 9 Literaturverzeichnis

- AAPOR. (2016). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (9th editio Aufl.). Autor.
- Biemer, P., Trewin, D., Bergdahl, H. & Japac, L. (2014, jan). A System for Managing the Quality of Official Statistics. *Journal of Official Statistics*, 30 (3), 381–415.
- Biemer, P. P. (2010, jan). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly*, 74 (5), 817–848.
- Callegaro, M., Manfreda, K. L. & Vehovar, V. (2015). *Web survey methodology*. Sage Publications.
- Committee, E. S. S. et al. (2012). *Quality assurance framework of the european statistical system*. Version.
- de Heij, V., Schouten, B. & Shlomo, N. (2010, Mai). Risq manual tools in sas and r for the computation of r-indicators and partial r-indicators work package 8 deliverable 12.1 [Software-Handbuch].
- Dillman, D., Smyth, J. & Christian, L. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. John Wiley & Sons.
- Dillman, D. A. & Edwards, M. L. (2016). *Designing a MixedMode Survey*. Sage.
- Ernst Stähli, M. & Joye, D. (2016). Incentives as a Possible Measure to Increase Response Rates. *The SAGE Handbook of Survey Methodology*.
- Göritz, A. S. (2006). Incentives in web studies: Methodological issues and a review. *International Journal of Internet Science*, 1 (1), 58–70.
- Groves, R. (2006). Nonresponse rates and nonresponse bias in household surveys: What Do We Know about the Linkage between Nonresponse Rates and Nonresponse Bias? *Public Opinion Quarterly*, 70 (5), 646–675.
- Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E. & Tourangeau, R. (2009). *Survey methodology*. John Wiley & Sons.
- Groves, R. M. & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly*, 74 (5), 849–879.
- Gumprecht, D. & Oismüller, A. (2013). Non-Response im Mikrozensus. *Statistische Nachrichten*, 11 (2013), 1046–1061.
- Kennedy, J. M. & Ouimet, J. A. (2014). The effect of incentives on internet surveys: Response rate changes after the introduction of incentives. *Asian Journal for Public Opinion Research*, 1 (2), 128–146.
- Kish, L. (1965). Survey sampling.

- Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, 9 (1), 1-19. (R package version 2.2)
- Lumley, T. (2016). *survey: analysis of complex survey samples*. (R package version 3.31-2)
- Lyberg, L. & Weisberg, H. (2016). *Total Survey Error: A Paradigm for Survey Methodology*. Sage Publications.
- Medway, R. L. & Tourangeau, R. (2015). Response quality in telephone surveys: Do prepaid cash incentives make a difference? *Public Opinion Quarterly*, 79 (2), 524–543.
- Meraner, A., Gumprecht, D. & Kowarik, A. (2016). Weighting Procedure of the Austrian Microcensus using Administrative Data. *Austrian Journal of Statistics*, 45 (3), 3.
- Porter, S. R., Whitcomb, M. E. & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*, 2004 (121), 63–73.
- Reinecke, J. (2014). Grundlagen der standardisierten Befragung. In *Handbuch methoden der empirischen sozialforschung* (S. 601–617). Wiesbaden: Springer Fachmedien Wiesbaden.
- Schouten, B., Cobben, F. & Bethlehem, J. (2009). Indicators for the Representativeness of Survey Response. *Survey Methodology*, 35 (1), 101–113.
- Simmons, E. & Wilmot, A. (2004). Incentive payments on social surveys: A literature review. *Social survey methodology bulletin*, 1–11.
- Weichbold, M., Bacher, J. & Wolf, C. (2009). *Umfrageforschung - Herausforderungen und Grenzen*.
- Weisberg, H. F. (2009). *The total survey error approach: A guide to the new science of survey research*. University of Chicago Press.

## **10 Anhang**

### **10.1 Kodierung der Variablen**



Tabelle 11: Dokumentation zur Umkodierung der ursprünglichen Variablen.

Name im Bericht	ursprüngliche Kodierung	Umkodierung für Kreuztabellen	Umkodierung für Analyse
Alter	Geburtsjahr	1 = 16 bis 24 Jahre,	1 = bis 24 Jahre,
		2 = 25 bis 34 Jahre,	2 = 25 bis 64 Jahre,
		3 = 35 bis 44 Jahre,	2 = 25 bis 64 Jahre,
		4 = 45 bis 54 Jahre,	2 = 25 bis 64 Jahre,
		5 = 55 bis 64 Jahre,	2 = 25 bis 64 Jahre,
		6 = 65 bis 74 Jahre	3 = 65 Jahre und mehr
Familienstand	1 = ledig	1 = ledig	1 = nicht verheiratet
	2 = verheiratet (oder eingetragene Partnerschaft)	2 = verheiratet (oder eingetragene Partnerschaft)	2 = verheiratet
	3 = verwitwet (oder hinterbliebene/r eingetragene/r Partner/in)	3 = verwitwet (oder hinterbliebene/r eingetragene/r Partner/in)	1 = nicht verheiratet
	4 = geschieden (oder aufgelöste eingetragene Partnerschaft)	4 = geschieden (oder aufgelöste eingetragene Partnerschaft)	1 = nicht verheiratet
Haushaltsgröße	1 = eine Person	1 = eine Person	1 = kleiner Haushalt
	:	:	1 = kleiner Haushalt
	5 = 5 Personen	5 = 5 Personen	2 = großer Haushalt (5+)
	6 = 6 Personen	6 = 6 Personen	2 = großer Haushalt (5+)
	7 = 7 Personen	7 = 7 und mehr	2 = großer Haushalt (5+)
	:	7 = 7 und mehr	2 = großer Haushalt (5+)

*Fortsetzung auf nächster Seite*

Tabelle 11 – Fortsetzung von vorheriger Seite

Name im Bericht	ursprüngliche Kodierung	Umkodierung für Kreuztabellen	Umkodierung für Analyse
	11 = mehr als 10 Personen	7 = 7 und mehr	2 = großer Haushalt (5+)
Bildungsabschluss	1 = kein Pflichtschulabschluss	1 = kein Pflichtschulabschluss	1 = kein Pflichtschulabschluss
	2 = Pflichtschulabschluss	2 = Pflichtschulabschluss	2 = Pflichtschulabschluss
	3 = Lehre (Lehrabschlussprüfung)	3 = Lehre (Lehrabschlussprüfung)	3 = Matura/Lehre
	4 = BMS (Fach- oder Handelsschule <2 Jahre und länger)	4 = BMS (Fach- oder Handelsschule <2 Jahre, 2 Jahre und länger)	3 = Matura/Lehre
	5 = AHS mit Matura	5 = AHS mit Matura	3 = Matura/Lehre
	6 = BHS mit Matura	6 = BHS mit Matura	3 = Matura/Lehre
	7 = Hochschulverwandte Lehristalt (Akademien)	7 = Hochschulverwandte Lehristalt (Akademien)	4 = Hochschule
	8 = Kolleg	8 = Kolleg	4 = Hochschule
	9 = Bachelor	9 = Bachelor	4 = Hochschule
	10 = Magister/Master/DI/FH	10 = Magister/Master/DI/FH	4 = Hochschule
	11 = Doktor/PhD	11 = Doktor/PhD	4 = Hochschule

Fortsetzung auf nächster Seite

Tabelle 11 – Fortsetzung von vorheriger Seite

Name im Bericht	ursprüngliche Kodierung	Umkodierung für Kreuztabellen	Umkodierung für Analyse
Hauptbeschäftigung	12 = Andere (= Diplomkran- kenpflege, BHS, 3. Klasse ab- geschlossen, Lehre mit Matu- ra, Postgradualer Universitäts- lehrgang, Hochschul- Universi- tätslehrgang (akademisch geprüf- te Berufsbezeichnung), Meister- oder Werkmeisterprüfung)	12 = Andere (= Diplomkran- kenpflege, BHS, 3. Klasse ab- geschlossen, Lehre mit Matu- ra, Postgradualer Universitäts- lehrgang, Hochschul- Universi- tätslehrgang (akademisch geprüf- te Berufsbezeichnung), Meister- oder Werkmeisterprüfung)	5 = Andere
	1 = in Pension	1 = in Pension	1 = in Pension
	2 = berufstätig	2 = berufstätig	2 = berufstätig
	3 = Schüler/Student	3 = Schüler/Student	3 = nicht berufstätig
	4 = in Elternkarenz	4 = in Elternkarenz	3 = nicht berufstätig
	5 = Hausfrau/Hausmann	5 = Hausfrau/Hausmann	3 = nicht berufstätig
	6 = Präsenzdienner/Zivildienner	NA = Missing <sup>20</sup>	3 = nicht berufstätig
	7 = berufsunfähig	7 = berufsunfähig	3 = nicht berufstätig
	8 = arbeitssuchend/arbeitslos	8 = arbeitssuchend/arbeitslos	3 = nicht berufstätig
	9 = Sonstiges	NA = Missing	3 = nicht berufstätig
	-3 = Filter	NA = Missing	NA = Missing
	NA = Missing	NA = Missing	Unbekannt <sup>21</sup>
Berufliche Stellung	1 = Angestellte/r	1 = Angestellte/r	1 = Angestellte/r
	2 = Arbeiter/in	2 = Arbeiter/in	2 = Arbeiter/in
	3 = Beamter/Beamtin	3 = Beamter/Beamtin	1 = Angestellte/r
	4 = Vertragsbedienstete/r	4 = Vertragsbedienstete/r	1 = Angestellte/r

Fortsetzung auf nächster Seite

<sup>20</sup>Präsenzdienner/Zivildienner wurden wegen zu kleinem n ausgeschlossen

<sup>21</sup>Dies betrifft 6313 Fälle und wird deshalb als Unbekannt mit aufgenommen.

Tabelle 11 – Fortsetzung von vorheriger Seite

Name im Bericht	ursprüngliche Kodierung	Umkodierung für Kreuztabellen	Umkodierung für Analyse
Österreichische Staatsbürgerschaft	5 = selbständig oder freiberuflich tätig ohne Mitarbeiter	5 = selbständig oder freiberuflich tätig ohne Mitarbeiter	3 = Selbständig/r
	6 = selbständig oder freiberuflich tätig mit Mitarbeitern	6 = selbständig oder freiberuflich tätig mit Mitarbeitern	3 = Selbständig/r
	7 = Freie/r Dienstnehmer/in	7 = Freie/r Dienstnehmer/in	3 = Selbständig/r
	8 = Landwirt	8 = Landwirt	3 = Selbständig/r
	1 = Ja	1 = Ja	1 = Ja
	2 = Nein	2 = Nein	2 = Nein
	1 = Österreich	1 = Österreich	1 = EU15 mit Österreich
	2 = EU 15 (ohne Österreich)	2 = EU 15 (ohne Österreich)	reich 1 = EU15 mit Österreich
Einkommen	3 = EU 25 (10 neuen ab 2004)	3 = EU 25 (10 neuen ab 2004)	2 = außerhalb EU15
	4 = Ex-Jugoslawien (ohne Slowenien, Kroatien)	4 = Ex-Jugoslawien (ohne Slowenien, Kroatien)	2 = außerhalb EU15
	5 = Türkei	5 = Türkei	2 = außerhalb EU15
	6 = Sonstige	6 = Sonstige	2 = außerhalb EU15
	7 = Bulgarien/Rumänien (EU27 ab 2007)	7 = Bulgarien/Rumänien (EU27 ab 2007)	2 = außerhalb EU15
	8 = Kroatien (EU28 ab Juli 2013)	8 = Kroatien (EU28 ab Juli 2013)	2 = außerhalb EU15
	Einkommensquintil 1 (niedrigstes) - 5 (höchstes)	Einkommensquintil 1 (niedrigstes) - 5 (höchstes)	Einkommensquintil 1 (niedrigstes) - 5 (höchstes)

## 10.2 Kreuztabellen zum Vergleich Teilstichprobe und Bruttostichprobe

Tabelle 12: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Incentivegruppen (ungewichtet)

Incentivegruppe	Bruttostichprobe		Gesamt
	Nein	Ja	
Broschüre	565 26,4%	1115 26,2%	1680 26,3%
2 EUR Münze	557 26,0%	1046 24,6%	1603 25,1%
5 EUR Münze	520 24,3%	1056 24,9%	1576 24,7%
Gutschein	497 23,2%	1032 24,3%	1529 23,9%
Gesamt	2139 100,0%	4249 100,0%	6388 100,0%

$$\chi^2 = 1,992; df = 3; p = 0,574$$

Tabelle 13: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Geschlecht (ungewichtet)

Geschlecht	Bruttostichprobe		Gesamt
	Nein	Ja	
Mann	1086 50,8%	2053 48,3%	3139 49,1%
Frau	1053 49,2%	2196 51,7%	3249 50,9%
Gesamt	2139 100,0%	4249 100,0%	6388 100,0%

$$\chi^2 = 3,429; df = 1; p = 0,064$$

Tabelle 14: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Altersgruppen (ungewichtet)

Altersgruppen	Bruttostichprobe		Gesamt
	Nein	Ja	
16 bis 24 Jahre	470 22,0%	331 7,8%	801 12,5%
25 bis 34 Jahre	297 13,9%	596 14,0%	893 14,0%
35 bis 44 Jahre	301 14,1%	812 19,1%	1113 17,4%
45 bis 54 Jahre	388 18,1%	1093 25,7%	1481 23,2%
55 bis 64 Jahre	350 16,4%	844 19,9%	1194 18,7%
65 bis 74 Jahre	333 15,6%	573 13,5%	906 14,2%
Gesamt	2139 100,0%	4249 100,0%	6388 100,0%

$$\chi^2 = 297,968; df = 5; p < 0,001$$

Tabelle 15: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Familienstand (ungewichtet)

Familienstand	Bruttostichprobe		Gesamt
	Nein	Ja	
Ledig	923 43,2%	1384 32,6%	2307 36,1%
Verheiratet (oder eingetragene Partnerschaft)	1007 47,1%	2248 52,9%	3255 51,0%
Verwitwet (oder hinterbliebene/r eingetragene/r Partner/in)	65 3,0%	150 3,5%	215 3,4%
Geschieden (oder aufgelöste eingetragene Partnerschaft)	144 6,7%	467 11,0%	611 9,6%
Gesamt	2139 100,0%	4249 100,0%	6388 100,0%

$$\chi^2 = 81,571; df = 3; p < 0,001$$

Tabelle 16: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Haushaltsgröße (ungewichtet)

Haushaltsgröße	Bruttostichprobe		Gesamt
	Nein	Ja	
1	233 10,9%	829 19,5%	1062 16,6%
2	702 32,8%	1436 33,8%	2138 33,5%
3	468 21,9%	846 19,9%	1314 20,6%
4	459 21,5%	808 19,0%	1267 19,8%
5	183 8,6%	248 5,8%	431 6,7%
6	71 3,3%	64 1,5%	135 2,1%
7 und mehr	23 1,1%	18 0,4%	41 0,6%
Gesamt	2139 100,0%	4249 100,0%	6388 100,0%

$$\chi^2 = 118,05; df = 6; p < 0,001$$

Tabelle 17: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Bildung (ungewichtet)

Bildung	Bruttostichprobe		Gesamt
	Nein	Ja	
kein Pflichtschulabschluss	46 2,2%	15 0,4%	61 1,0%
Pflichtschulabschluss	590 27,6%	544 12,8%	1134 17,8%
Lehre (Lehrabschlussprüfung)	751 35,1%	1565 36,8%	2316 36,3%
BMS (Fach- oder handelsschule <2 Jahre, 2 Jahre und länger)	202 9,4%	465 10,9%	667 10,4%
AHS mit Matura	114 5,3%	222 5,2%	336 5,3%
BHS mit Matura	135 6,3%	428 10,1%	563 8,8%
Hochschulverwandte Lehranstalt (Akademien)	21 1,0%	87 2,0%	108 1,7%
Kolleg	2 0,1%	19 0,4%	21 0,3%
Bachelor	41 1,9%	88 2,1%	129 2,0%
Magister/Master/DI/FH	78 3,6%	362 8,5%	440 6,9%
Doktor/PhD	27 1,3%	91 2,1%	118 1,8%
Andere	132 6,2%	363 8,5%	495 7,7%
Gesamt	2139 100,0%	4249 100,0%	6388 100,0%

$\chi^2 = 330,8; df = 11; p < 0,001$



Tabelle 18: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Lebensunterhalt (ungewichtet)

Lebensunterhalt	Bruttostichprobe		Gesamt
	Nein	Ja	
in Pension	448 21,2%	846 20,0%	1294 20,4%
berufstätig	1112 52,6%	2780 65,6%	3892 61,2%
Schüler/Student	265 12,5%	178 4,2%	443 7,0%
in Elternkarenz	41 1,9%	66 1,6%	107 1,7%
Hausfrau/Hausmann	105 5,0%	165 3,9%	270 4,2%
berufsunfähig	50 2,4%	50 1,2%	100 1,6%
arbeitssuchend/arbeitslos	95 4,5%	154 3,6%	249 3,9%
Gesamt	2116 100,0%	4239 100,0%	6355 100,0%

$$\chi^2 = 200,682; df = 6; p < 0,001$$

Tabelle 19: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach beruflicher Stellung (ungewichtet)

Berufliche Stellung	Bruttostichprobe		Gesamt
	Nein	Ja	
Angestellte/r	496 42,1%	1563 52,5%	2059 49,5%
Arbeiter/in	458 38,9%	647 21,7%	1105 26,6%
Beamter/Beamtin	34 2,9%	191 6,4%	225 5,4%
Vertragsbedienstete	52 4,4%	177 5,9%	229 5,5%
selbständig oder freiberuflich tätig ohne Mitarbeiter	52 4,4%	167 5,6%	219 5,3%
selbständig oder freiberuflich tätig mit Mitarbeitern	56 4,8%	144 4,8%	200 4,8%
freie/r Dienstnehmer/in	8 0,7%	23 0,8%	31 0,7%
Landwirt	21 1,8%	67 2,2%	88 2,1%
Gesamt	1177 100,0%	2979 100,0%	4156 100,0%

$$\chi^2 = 138,085; df = 7; p < 0,001$$

Tabelle 20: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Staatsangehörigkeit (ungewichtet)

Staatsangehörigkeit	Bruttostichprobe		Gesamt
	Nein	Ja	
Ja	1784 83,4%	3904 91,9%	5688 89,0%
Nein	355 16,6%	345 8,1%	700 11,0%
Gesamt	2139 100,0%	4249 100,0%	6388 100,0%

$\chi^2 = 104,782; df = 1; p < 0,001$

Tabelle 21: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Herkunft (ungewichtet)

Herkunft	Bruttostichprobe		Gesamt
	Nein	Ja	
Österreich	1671 78,1%	3683 86,7%	5354 83,8%
EU 15 (ohne Ö)	56 2,6%	173 4,1%	229 3,6%
EU25 (10 neuen ab 2004)	53 2,5%	73 1,7%	126 2,0%
Ex-Jugoslawien (ohne Slowenien, Kroatien)	146 6,8%	136 3,2%	282 4,4%
Türkei	82 3,8%	60 1,4%	142 2,2%
Sonstige	82 3,8%	77 1,8%	159 2,5%
Bulgarien/Rumänien (EU27 ab 2007)	40 1,9%	36 0,8%	76 1,2%
Kroatien (EU28 ab Juli 2013)	9 0,4%	11 0,3%	20 0,3%
Gesamt	2139 100,0%	4249 100,0%	6388 100,0%

$\chi^2 = 141,916; df = 7; p < 0,001$

Tabelle 22: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Bruttostichprobe nach Einkommensquintilen (ungewichtet)

Einkommen	Bruttostichprobe		Gesamt
	Nein	Ja	
1	369 17,3%	598 14,1%	967 15,1%
2	436 20,4%	746 17,6%	1182 18,5%
3	429 20,1%	796 18,7%	1225 19,2%
4	485 22,7%	982 23,1%	1467 23,0%
5	420 19,6%	1127 26,5%	1547 24,2%
Gesamt	2139 100,0%	4249 100,0%	6388 100,0%

$$\chi^2 = 44,923; df = 4; p < 0,001$$

### 10.3 Kreuztabellen zum Vergleich Teilstichprobe und Nettostichprobe

Tabelle 23: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Incentivegruppen (ungewichtet)

Incentivegruppe	Nettostichprobe		Gesamt
	Nein	Ja	
Broschüre	1490 27,9%	190 18,2%	1680 26,3%
2 EUR Münze	1327 24,8%	276 26,5%	1603 25,1%
5 EUR Münze	1282 24,0%	294 28,2%	1576 24,7%
Gutschein	1246 23,3%	283 27,1%	1529 23,9%
Gesamt	5345 100,0%	1043 100,0%	6388 100,0%

$$\chi^2 = 43,468; df = 3; p < 0,001$$

Tabelle 24: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Geschlecht (ungewichtet)

Geschlecht	Nettostichprobe		Gesamt
	Nein	Ja	
Mann	2622 49,1%	517 49,6%	3139 49,1%
Frau	2723 50,9%	526 50,4%	3249 50,9%
Gesamt	5345 100,0%	1043 100,0%	6388 100,0%

$$\chi^2 = 0,092; df = 1; p = 0,762$$

Tabelle 25: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Altersgruppen (ungewichtet)

Altersgruppen	Nettostichprobe		Gesamt
	Nein	Ja	
16 bis 24 Jahre	730 13,7%	71 6,8%	801 12,5%
25 bis 34 Jahre	735 13,8%	158 15,1%	893 14,0%
35 bis 44 Jahre	864 16,2%	249 23,9%	1113 17,4%
45 bis 54 Jahre	1173 21,9%	308 29,5%	1481 23,2%
55 bis 64 Jahre	1024 19,2%	170 16,3%	1194 18,7%
65 bis 74 Jahre	819 15,3%	87 8,3%	906 14,2%
Gesamt	5345 100,0%	1043 100,0%	6388 100,0%

$$\chi^2 = 119,106; df = 5; p < 0,001$$

Tabelle 26: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Familienstand (ungewichtet)

Familienstand	Nettostichprobe		Gesamt
	Nein	Ja	
Ledig	1950 36,5%	357 34,2%	2307 36,1%
Verheiratet (oder eingetragene Partnerschaft)	2671 50,0%	584 56,0%	3255 51,0%
Verwitwet (oder hinterbliebene/r eingetragene/r Partner/in)	202 3,8%	13 1,2%	215 3,4%
Geschieden (oder aufgelöste eingetragene Partnerschaft)	522 9,8%	89 8,5%	611 9,6%
Gesamt	5345 100,0%	1043 100,0%	6388 100,0%

$$\chi^2 = 25,458; df = 3; p < 0,001$$

Tabelle 27: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Haushaltsgröße (ungewichtet)

Haushaltsgröße	Nettostichprobe		Gesamt
	Nein	Ja	
1	890 16,7%	172 16,5%	1062 16,6%
2	1778 33,3%	360 34,5%	2138 33,5%
3	1095 20,5%	219 21,0%	1314 20,6%
4	1047 19,6%	220 21,1%	1267 19,8%
5	378 7,1%	53 5,1%	431 6,7%
6	120 2,2%	15 1,4%	135 2,1%
7 und mehr	37 0,7%	4 0,4%	41 0,6%
Gesamt	5345 100,0%	1043 100,0%	6388 100,0%

$$\chi^2 = 10,638; df = 6; p = 0,1$$

Tabelle 28: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Bildung (ungewichtet)

Bildung	Nettostichprobe		Gesamt
	Nein	Ja	
kein Pflichtschulabschluss	61 1,1%	0 0,0%	61 1,0%
Pflichtschulabschluss	1099 20,6%	35 3,4%	1134 17,8%
Lehre (Lehrabschlussprüfung)	2044 38,2%	272 26,1%	2316 36,3%
BMS (Fach- oder handelsschule <2 Jahre, 2 Jahre und länger)	563 10,5%	104 10,0%	667 10,4%
AHS mit Matura	258 4,8%	78 7,5%	336 5,3%
BHS mit Matura	406 7,6%	157 15,1%	563 8,8%
Hochschulverwandte Lehranstalt (Akademien)	69 1,3%	39 3,7%	108 1,7%
Kolleg	13 0,2%	8 0,8%	21 0,3%
Bachelor	89 1,7%	40 3,8%	129 2,0%
Magister/Master/DI/FH	282 5,3%	158 15,1%	440 6,9%
Doktor/PhD	66 1,2%	52 5,0%	118 1,8%
Andere	395 7,4%	100 9,6%	495 7,7%
Gesamt	5345 100,0%	1043 100,0%	6388 100,0%

$\chi^2 = 514,006; df = 11; p < 0,001$

Tabelle 29: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Lebensunterhalt (ungewichtet)

Lebensunterhalt	Nettostichprobe		Gesamt
	Nein	Ja	
in Pension	1161 21,8%	133 12,8%	1294 20,4%
berufstätig	3121 58,7%	771 74,2%	3892 61,2%
Schüler/Student	384 7,2%	59 5,7%	443 7,0%
in Elternkarenz	85 1,6%	22 2,1%	107 1,7%
Hausfrau/Hausmann	250 4,7%	20 1,9%	270 4,2%
berufsunfähig	99 1,9%	1 0,1%	100 1,6%
arbeitssuchend/arbeitslos	216 4,1%	33 3,2%	249 3,9%
Gesamt	5316 100,0%	1039 100,0%	6355 100,0%

$$\chi^2 = 108,075; df = 6; p < 0,001$$

Tabelle 30: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach beruflicher Stellung (ungewichtet)

Berufliche Stellung	Nettostichprobe		Gesamt
	Nein	Ja	
Angestellte/r	1554 46,8%	505 60,4%	2059 49,5%
Arbeiter/in	1027 30,9%	78 9,3%	1105 26,6%
Beamter/Beamtin	138 4,2%	87 10,4%	225 5,4%
Vertragsbedienstete	163 4,9%	66 7,9%	229 5,5%
selbständig oder freiberuflich tätig ohne Mitarbeiter	163 4,9%	56 6,7%	219 5,3%
selbständig oder freiberuflich tätig mit Mitarbeitern	176 5,3%	24 2,9%	200 4,8%
freie/r Dienstnehmer/in	21 0,6%	10 1,2%	31 0,7%
Landwirt	78 2,3%	10 1,2%	88 2,1%
Gesamt	3320 100,0%	836 100,0%	4156 100,0%

$$\chi^2 = 220,439; df = 7; p < 0,001$$



Tabelle 31: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Staatsangehörigkeit (ungewichtet)

Staatsangehörigkeit	Nettostichprobe		Gesamt
	Nein	Ja	
Ja	4707 88,1%	981 94,1%	5688 89,0%
Nein	638 11,9%	62 5,9%	700 11,0%
Gesamt	5345 100,0%	1043 100,0%	6388 100,0%

$\chi^2 = 32,113; df = 1; p < 0,001$

Tabelle 32: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Herkunft (ungewichtet)

Herkunft	Nettostichprobe		Gesamt
	Nein	Ja	
Österreich	4406 82,4%	948 90,9%	5354 83,8%
EU 15 (ohne Ö)	176 3,3%	53 5,1%	229 3,6%
EU25 (10 neuen ab 2004)	110 2,1%	16 1,5%	126 2,0%
Ex-Jugoslawien (ohne Slowenien, Kroatien)	274 5,1%	8 0,8%	282 4,4%
Türkei	138 2,6%	4 0,4%	142 2,2%
Sonstige	150 2,8%	9 0,9%	159 2,5%
Bulgarien/Rumänien (EU27 ab 2007)	74 1,4%	2 0,2%	76 1,2%
Kroatien (EU28 ab Juli 2013)	17 0,3%	3 0,3%	20 0,3%
Gesamt	5345 100,0%	1043 100,0%	6388 100,0%

$\chi^2 = 96,699; df = 7; p < 0,001$

Tabelle 33: Vergleich Teilstichprobe Mikrozensus (Rotation 50) und Nettostichprobe nach Einkommensquintilen (ungewichtet)

Einkommen	Nettostichprobe		Gesamt
	Nein	Ja	
1	852 15,9%	115 11,0%	967 15,1%
2	1041 19,5%	141 13,5%	1182 18,5%
3	1044 19,5%	181 17,4%	1225 19,2%
4	1238 23,2%	229 22,0%	1467 23,0%
5	1170 21,9%	377 36,1%	1547 24,2%
Gesamt	5345 100,0%	1043 100,0%	6388 100,0%

$$\chi^2 = 106,613; df = 4; p < 0,001$$

## 10.4 Kreuztabellen zum Vergleich Brutto- und Nettostichprobe

Tabelle 34: Vergleich Brutto- und Nettostichprobe nach Incentivegruppen (ungewichtet)

Incentivegruppe	Nettostichprobe		Gesamt
	Nein	Ja	
Broschüre	925 28,9%	190 18,2%	1115 26,2%
2 EUR Münze	770 24,0%	276 26,5%	1046 24,6%
5 EUR Münze	762 23,8%	294 28,2%	1056 24,9%
Gutschein	749 23,4%	283 27,1%	1032 24,3%
Gesamt	3206 100,0%	1043 100,0%	4249 100,0%

$$\chi^2 = 46,626; df = 3; p < 0,001$$

,

Tabelle 35: Vergleich Brutto- und Nettostichprobe nach Geschlecht (ungewichtet)

Geschlecht	Nettostichprobe		Gesamt
	Nein	Ja	
Mann	1536 47,9%	517 49,6%	2053 48,3%
Frau	1670 52,1%	526 50,4%	2196 51,7%
Gesamt	3206 100,0%	1043 100,0%	4249 100,0%

$$\chi^2 = 0,867; df = 1; p = 0,352$$

Tabelle 36: Vergleich Brutto- und Nettostichprobe nach Altersgruppen (ungewichtet)

Altersgruppen	Nettostichprobe		Gesamt
	Nein	Ja	
16 bis 24 Jahre	260 8,1%	71 6,8%	331 7,8%
25 bis 34 Jahre	438 13,7%	158 15,1%	596 14,0%
35 bis 44 Jahre	563 17,6%	249 23,9%	812 19,1%
45 bis 54 Jahre	785 24,5%	308 29,5%	1093 25,7%
55 bis 64 Jahre	674 21,0%	170 16,3%	844 19,9%
65 bis 74 Jahre	486 15,2%	87 8,3%	573 13,5%
Gesamt	3206 100,0%	1043 100,0%	4249 100,0%

$$\chi^2 = 63,117; df = 5; p < 0,001$$

Tabelle 37: Vergleich Brutto- und Nettostichprobe nach Familienstand (ungewichtet)

Familienstand	Nettostichprobe		Gesamt
	Nein	Ja	
Ledig	1027 32,0%	357 34,2%	1384 32,6%
Verheiratet (oder eingetragene Partnerschaft)	1664 51,9%	584 56,0%	2248 52,9%
Verwitwet (oder hinterbliebene/r eingetragene/r Partner/in)	137 4,3%	13 1,2%	150 3,5%
Geschieden (oder aufgelöste eingetragene Partnerschaft)	378 11,8%	89 8,5%	467 11,0%
Gesamt	3206 100,0%	1043 100,0%	4249 100,0%

$$\chi^2 = 31,672; df = 3; p < 0,001$$

Tabelle 38: Vergleich Brutto- und Nettostichprobe nach Haushaltsgröße (ungewichtet)

Haushaltsgröße	Nettostichprobe		Gesamt
	Nein	Ja	
1	657 20,5%	172 16,5%	829 19,5%
2	1076 33,6%	360 34,5%	1436 33,8%
3	627 19,6%	219 21,0%	846 19,9%
4	588 18,3%	220 21,1%	808 19,0%
5	195 6,1%	53 5,1%	248 5,8%
6	49 1,5%	15 1,4%	64 1,5%
7 und mehr	14 0,4%	4 0,4%	18 0,4%
Gesamt	3206 100,0%	1043 100,0%	4249 100,0%

$$\chi^2 = 12,072; df = 6; p = 0,06$$

Tabelle 39: Vergleich Brutto- und Nettostichprobe nach Bildung (ungewichtet)

Bildung	Nettostichprobe		Gesamt
	Nein	Ja	
kein Pflichtschulabschluss	15 0,5%	0 0,0%	15 0,4%
Pflichtschulabschluss	509 15,9%	35 3,4%	544 12,8%
Lehre (Lehrabschlussprüfung)	1293 40,3%	272 26,1%	1565 36,8%
BMS (Fach- oder handelsschule <2 Jahre, 2 Jahre und länger)	361 11,3%	104 10,0%	465 10,9%
AHS mit Matura	144 4,5%	78 7,5%	222 5,2%
BHS mit Matura	271 8,5%	157 15,1%	428 10,1%
Hochschulverwandte Lehranstalt (Akademien)	48 1,5%	39 3,7%	87 2,0%
Kolleg	11 0,3%	8 0,8%	19 0,4%
Bachelor	48 1,5%	40 3,8%	88 2,1%
Magister/Master/DI/FH	204 6,4%	158 15,1%	362 8,5%
Doktor/PhD	39 1,2%	52 5,0%	91 2,1%
Andere	263 8,2%	100 9,6%	363 8,5%
Gesamt	3206 100,0%	1043 100,0%	4249 100,0%

$\chi^2 = 361,823; df = 11; p < 0,001$

Tabelle 40: Vergleich Brutto- und Nettostichprobe nach Lebensunterhalt (ungewichtet)

Lebensunterhalt	Nettostichprobe		Gesamt
	Nein	Ja	
in Pension	713 22,3%	133 12,8%	846 20,0%
berufstätig	2009 62,8%	771 74,2%	2780 65,6%
Schüler/Student	119 3,7%	59 5,7%	178 4,2%
in Elternkarenz	44 1,4%	22 2,1%	66 1,6%
Hausfrau/Hausmann	145 4,5%	20 1,9%	165 3,9%
berufsunfähig	49 1,5%	1 0,1%	50 1,2%
arbeitssuchend/arbeitslos	121 3,8%	33 3,2%	154 3,6%
Gesamt	3200 100,0%	1039 100,0%	4239 100,0%

$$\chi^2 = 89,055; df = 6; p < 0,001$$

Tabelle 41: Vergleich Brutto- und Nettostichprobe nach beruflicher Stellung (ungewichtet)

Berufliche Stellung	Nettostichprobe		Gesamt
	Nein	Ja	
Angestellte/r	1058 49,4%	505 60,4%	1563 52,5%
Arbeiter/in	569 26,6%	78 9,3%	647 21,7%
Beamter/Beamtin	104 4,9%	87 10,4%	191 6,4%
Vertragsbedienstete	111 5,2%	66 7,9%	177 5,9%
selbständig oder freiberuflich tätig ohne Mitarbeiter	111 5,2%	56 6,7%	167 5,6%
selbständig oder freiberuflich tätig mit Mitarbeitern	120 5,6%	24 2,9%	144 4,8%
freie/r Dienstnehmer/in	13 0,6%	10 1,2%	23 0,8%
Landwirt	57 2,7%	10 1,2%	67 2,2%
Gesamt	2143 100,0%	836 100,0%	2979 100,0%

$$\chi^2 = 152,651; df = 7; p < 0,001$$

Tabelle 42: Vergleich Brutto- und Nettostichprobe nach Staatsangehörigkeit (ungewichtet)

Staatsangehörigkeit	Nettostichprobe		Gesamt
	Nein	Ja	
Ja	2923 91,2%	981 94,1%	3904 91,9%
Nein	283 8,8%	62 5,9%	345 8,1%
Gesamt	3206 100,0%	1043 100,0%	4249 100,0%

$\chi^2 = 8,767; df = 1; p = 0,003$

Tabelle 43: Vergleich Brutto- und Nettostichprobe nach Herkunft (ungewichtet)

Herkunft	Nettostichprobe		Gesamt
	Nein	Ja	
Österreich	2735 85,3%	948 90,9%	3683 86,7%
EU 15 (ohne Ö)	120 3,7%	53 5,1%	173 4,1%
EU25 (10 neuen ab 2004)	57 1,8%	16 1,5%	73 1,7%
Ex-Jugoslawien (ohne Slowenien, Kroatien)	128 4,0%	8 0,8%	136 3,2%
Türkei	56 1,7%	4 0,4%	60 1,4%
Sonstige	68 2,1%	9 0,9%	77 1,8%
Bulgarien/Rumänien (EU27 ab 2007)	34 1,1%	2 0,2%	36 0,8%
Kroatien (EU28 ab Juli 2013)	8 0,2%	3 0,3%	11 0,3%
Gesamt	3206 100,0%	1043 100,0%	4249 100,0%

$\chi^2 = 56,431; df = 7; p < 0,001$



Tabelle 44: Vergleich Brutto- und Nettostichprobe nach Einkommensquintilen (unge-  
wichtet)

Einkommen	Nettostichprobe		Gesamt
	Nein	Ja	
1	483 15,1%	115 11,0%	598 14,1%
2	605 18,9%	141 13,5%	746 17,6%
3	615 19,2%	181 17,4%	796 18,7%
4	753 23,5%	229 22,0%	982 23,1%
5	750 23,4%	377 36,1%	1127 26,5%
Gesamt	3206 100,0%	1043 100,0%	4249 100,0%

$$\chi^2 = 72,418; df = 4; p < 0,001$$



## 10.5 Ergebnisse logistische Regression

Tabelle 45: Logistische Regressionskoeffizienten der Stichprobenausschöpfung (von Mikro- zu Teil-, von Teil- zu Brutto- und von Teil- zu Nettostichprobe, Standardfehler in Klammern).

PUMA-Variable	Variable in Regression	Chance der Selektion		
		Mik -> Tei <i>logistic</i>	Tei -> Bru <i>logistic</i>	Tei -> Net <i>logistic</i>
	Konstante	-1,538*** (0,089)	0,925*** (0,180)	-1,778*** (0,255)
Alter (Ref = 25 bis 64 Jahre)	65 Jahre und mehr	0,054 (0,059)	-0,233** (0,113)	-0,249 (0,168)
	bis 24 Jahre	-0,004 (0,050)	-0,984*** (0,096)	-0,224 (0,150)
In Partnerschaft (Ref = Nein)	Ja	-0,044 (0,031)	-0,017 (0,063)	0,219*** (0,077)
Haushaltsgröße (Ref = Haushalt (5+))	Haushalt (1-4 Personen)	0,074 (0,049)	0,322*** (0,095)	0,268* (0,140)
Bildung (Ref = Hochschulabschluss)	Kein Pflichtschulabschluss	0,193 (0,152)	-1,495*** (0,321)	-13,835 (181,929)
	Matura/Lehre	0,091** (0,044)	-0,329*** (0,099)	-0,899*** (0,092)
	Pflichtschulabschluss	0,083 (0,056)	-0,748*** (0,118)	-2,247*** (0,198)
	Anderer Abschluss	0,139** (0,064)	-0,138 (0,139)	-0,686*** (0,138)
Hauptbeschäftigung (Ref = berufstätig)	In Pension	0,095 (0,081)	0,404** (0,166)	0,098 (0,210)
	nicht berufstätig	0,031 (0,066)	0,150 (0,138)	0,294* (0,155)
Berufliche Stellung (Ref = Angestellte/r)	Arbeiter/in	0,039 (0,042)	-0,537*** (0,084)	-0,997*** (0,131)
	Selbständige/r	0,071 (0,054)	-0,210* (0,113)	-0,470*** (0,125)
	Unbekannte/nicht zutreffend	-0,085 (0,069)	-0,834*** (0,144)	-0,856*** (0,172)
Österreichische Staatsbürgerschaft (Ref = Ja)	Nein	-0,147*** (0,056)	-0,378*** (0,110)	-0,169 (0,165)
Herkunft (Ref = nicht EU15)	EU15 mit Österreich	-0,096* (0,055)	0,512*** (0,105)	1,025*** (0,188)
Einkommen	Einkommensquintile (5 = höchstes)	0,020* (0,011)	-0,019 (0,022)	0,034 (0,028)
Pseudo-R <sup>2</sup>		0,001	0,081	0,115
N		33.223	6.388	6.388
Log Likelihood		-16.246,910	-3.741,177	-2.517,320
Akaike Inf. Crit.		32.527,810	7.516,354	5.068,640

Anmerkung:

Tabelle 46: Logistische Regressionskoeffizienten der Stichprobenausschöpfung von Brutto- zu Nettostichprobe nach Incentivegruppen (Standardfehler in Klammern).

		Chance der Selektion			
		Broschüre	2 EUR Münze	5 EUR Münze	Gutschein
		<i>logistic</i>	<i>logistic</i>	<i>logistic</i>	<i>logistic</i>
	Konstante	-2,279*** (0,607)	-1,463*** (0,563)	-0,654 (0,526)	-0,749 (0,499)
Alter (Ref = 25 bis 64 Jahre)	65 Jahre und mehr	-0,134 (0,385)	0,115 (0,344)	-0,134 (0,353)	-0,735** (0,350)
	bis 24 Jahre	0,127 (0,354)	-0,206 (0,349)	0,087 (0,299)	0,646** (0,311)
In Partnerschaft (Ref = Nein)	Ja	0,251 (0,181)	0,001 (0,159)	0,324** (0,158)	0,363** (0,163)
Haushaltsgröße (Ref = Haushalt (5+))	Haushalt (1-4 Personen)	0,401 (0,324)	0,073 (0,312)	-0,020 (0,285)	0,194 (0,307)
Bildung (Ref = Hochschulabschluss)	Kein Pflichtschulabschluss	-12,822 (380,613)	-11,984 (377,838)	-11,916 (535,411)	-14,480 (544,104)
	Matura/Lehre	-0,872*** (0,212)	-0,875*** (0,189)	-0,864*** (0,199)	-0,903*** (0,198)
	Pflichtschulabschluss	-2,921*** (0,751)	-2,197*** (0,414)	-2,053*** (0,352)	-1,909*** (0,369)
	Anderer Abschluss	-0,499 (0,307)	-0,466 (0,285)	-0,884*** (0,307)	-0,900*** (0,293)
Hauptbeschäftigung (Ref = berufstätig)	In Pension	0,971** (0,467)	-0,357 (0,456)	-0,911** (0,457)	0,182 (0,425)
	nicht berufstätig	0,782** (0,360)	0,223 (0,339)	-0,231 (0,339)	0,335 (0,314)
Berufliche Stellung (Ref = Angestellte/r)	Arbeiter/in	-1,521*** (0,439)	-0,458* (0,243)	-0,988*** (0,265)	-1,038*** (0,256)
	Selbständige/r	-0,325 (0,286)	-0,496* (0,259)	-0,344 (0,253)	-0,610** (0,265)
	Unbekannte/nicht zutreffend	-1,188*** (0,381)	-0,403 (0,386)	-0,090 (0,366)	-0,806** (0,355)
Österreichische Staatsbürgerschaft (Ref = Ja)	Nein	0,072 (0,371)	0,228 (0,348)	0,021 (0,337)	-0,320 (0,364)
Herkunft (Ref = nicht EU15)	EU15 mit Österreich	0,622 (0,435)	1,106*** (0,392)	0,816** (0,393)	0,712* (0,374)
Einkommen	Einkommensquintile (5 = höchstes)	0,163** (0,068)	0,096* (0,057)	0,005 (0,055)	-0,020 (0,059)
Pseudo-R <sup>2</sup>		0,112	0,089	0,098	0,103
N		1.115	1.046	1.056	1.032
Log Likelihood		-452,049	-549,793	-563,084	-543,668
Akaike Inf. Crit.		938,097	1.133,586	1.160,168	1.121,335

Anmerkung:

\*p<0,1; \*\*p<0,05; \*\*\*p<0,01